Dependence in macroeconomic variables: Assessing instantaneous and persistent relations between and within time series

Dissertation zur Erlangung des Doktorgrades
der Wirtschaftswissenschaftlichen Fakultät
der Georg-August-Universität Göttingen

vorgelegt von
Simone Maxand
geboren in Bad Honnef

Göttingen, 2017

Acknowledgments

First of all, I thank Helmut Herwartz for supporting me along the whole avenue of my doctoral studies. I am very thankful that he did not grew tired in convincing me that sentences between formulas can sound elegant instead of bizarre. I thank him for his great ideas, for understanding my chaotic drafts of mathematical inaccuracies and for his support in ordering my thoughts. Without his enthusiasm, creativity and skepticism this thesis would be far away from its present form.

I thank Thomas Kneib for taking the role as second reviewer and mentoring me throughout the past years. Hopefully, he will at least partly admit some usefulness to time series analysis after reading my thesis. I want to thank him and all the other 'early' birds for the regular coffee break shortening the long way to work. Along these lines, I thank the whole (perhaps) world-best working group allowing me to spend now almost four years in a very productive and beautiful surrounding. I thank them for their sense of humor, pleasant moments between and after work and enabling such a nice time in the beautiful statoek-villa. I am very grateful to my colleagues and the alumni for being inspiring and supportive throughout. Especially, I thank my double office colleagues for the caring working atmosphere and for sharing all desirable office facilities. Furthermore, I thank Alexander Silbersdorff, Peter Pütz, Maren Ulm, Hannes Rohloff, Katharina Schley, Gábor Uhrin, Ciara Higgins and Johannes Pelda for careful proofreading. Related to this, I thank them, and Svenja Strauß, for their motivation during the final working phase.

I thank all my former and present Göttingen flat mates and friends for mimicking everything you might need, being almost family and best friends; for sharing delicious meals, home, enlightening discussions, pure fun, vacations, extensive parties and relaxing weekends; for feeling so comfortable in Göttingen and making it almost impossible for me to leave this city. Last but not least, I thank my parents and my whole family for building this solid foundation for my doctoral studies and being understanding and supportive in all possible directions.

Abstract

The present thesis comprises two rather independent chapters. In general, the diagnosis and quantification of dependence is a major aim of econometric studies. Along these lines, the concept of dependence serves as an encompassing framework to analyze time series with two very different techniques.

First, we consider a single macroeconomic time series. A series which incorporates only temporary deviations from deterministic terms provides a different starting point for economic interpretations than an 'unpredictable' (random) series. In the context of dependency, we are interested if a time series is stationary or if it exhibits persistent dependence on larger horizons. We consider univariate tests for stationarity under distinct model settings. More recently, panel unit root tests have been developed to overcome power deficiencies and provide a more general economic statement. The standard panel unit root tests are not robust under time-varying variances and trending data. Against this background, we introduce a new test procedure which performs well in this setting.

Departing from the framework of a single time series the diagnosis of dependencies between several variables provides evidence on relations in a macroeconomic system. Assuming stationarity of the series, we analyze instantaneous and persistent effects between macroeconomic indices by means of vector autoregressive models. Dependencies can, beyond the standard linear setting, be present in diverse forms. We first refer to the variety of dependencies. Subsequently, we review and compare nonparametric measures which are developed to robustly diagnose various dependence types. Drawing on these dependence diagnostics helps to conduct a preliminary analysis of the data. In macroeconomics, the analyst might be further interested in causalities. We analyze causalities by means of structural vector autoregressive models tracing the variables back to unanticipated independent shocks. Hereby, we are mainly interested in the identification of instantaneous response matrices relying on non-Gaussianity of the structural shocks. We compare such independence based identification procedures relying on beforehand selected nonparametric dependence measures. Furthermore, we highlight their performance by means of a simulation study. The assumption of at most one Gaussian (structural) shocks is essential for unique identification of the instantaneous response to independent shocks. However, in a system with multiple Gaussian shocks the non-Gaussian ones can still be uniquely identified. We prove this result and, thereby, enable a more general definition of identifiability.

Contents

1 Introduction

Since the very first confrontation of religion and science there has been an on-going debate on whether events are random or deterministic. The belief in the determinacy of the world implies that causes and effects can be known, i.e. that events are dependent in a temporal succession. In a cross-individual dimension, *dependencies* might indicate immediate relations between events or agents. At the same time, linguistically the term dependence often implicitly specifies its direction, e.g., 'profit depends on the production'. While one could get lost in thoughts and discussions on the existence and exact definition of (in)dependence, we focus on a rather straightforward concept by drawing on statistical terminology. Refraining from the assumption that causes and effects are determined and can be clearly separated allows for the indeterminacy of events. Speculations about existing structures and rising complexity within a system renders the exact determination of relations impossible. This gives rise to the disciplines of statistics and econometrics which evaluate whether dependencies are rather likely or not. In the following, dependence denotes an association between random variables, where the direction of dependence is not necessarily determined. We prefer to use the term *causality* for directed dependence.

The present thesis comprises two rather independent chapters. In general, the diagnosis and quantification of dependence is a major aim of econometric studies. Along these lines, the concept of dependence serves as an encompassing framework to analyze time series with two very different techniques. While approaching dependence in statistical terms, this concept has also high significance in economics. Setting economic variables in the wider context of a whole macroeconomic system might help to preview the consequences and the sources of decision making and to evaluate the relationships suggested in economic theory. In the following, the techniques to assess dependence in distinct dimensions of macroeconomic systems relies on time series models. A multivariate macroeconomic time series allows for the analysis of dependence in two dimensions, along the time axis and across variables. Dependence in the temporal dimension of one variable indicates how persistent the effects of unanticipated changes are. Analyzing the independence of the series' own past is informative, for instance, about the predictability of a time series. Within a macroeconomic system the definition of cross-sectional dependence or even causality might not be straightforward. Contemplating the contrary, i.e. independence, statistical methods allow for tracing the macroeconomic indices back to independent sources. This enables the interpretation of their effects separately and, thereby, the identification of causalities. The present dynamic econometric model links one or more macroeconomic variables linearly, conditional on their lags in previous time periods. The following section introduces the main model to gain the first insight into the data-generating process used throughout.

1.1 The dynamic econometric model

A univariate time series model allows the study of dynamic impact relations within a time series itself. More generally, a multivariate K-dimensional model additionally captures dynamic effects between variables. The standard *vector autoregressive (VAR) model* reads as

$$y_t = c_t + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \quad t = 1, \ldots, T, \tag{1.1}$$

with deterministic terms c_t, the K-dimensional vector y_t and $K \times K$ coefficient matrices A_i, $i = 1, \ldots, p$. The vector u_t captures the random terms. Throughout the thesis, a model formulation similar to Equation (1.1) describes the data-generating process theoretically, where we define the exact formulation separately in each chapter. For an observed data sample the theoretical process is, however, not known (or might not even exist). Consequently, for further statistical evaluation an analyst mostly applies a model which describes the given data best. In the following, we assume that standard techniques for model selection (e.g. lag order choice) work properly, and we derive the most appropriate (linear) model. It is needless to say that this specific model formulation excludes a wide set of alternative representations. However, we do not claim to completely cover distinct relations on the temporal axis and across variables but rather exploit the described formulations, focusing on unit root tests (Chapter 2) and structural analyses (Chapter 3). In the setting of multivariate time series, alternative model formulations include, for instance, vector autoregressive moving average (VARMA) models and vector error correction models (VECM). Nevertheless, in the following we presume prior transformations to a VAR model and either assume stationarity of the variables or, if integrated, no cointegration. In particular, parts of Chapter 3 can be further extended to underlying processes of nonlinear form. Details on alternative model formulations can be found in, for instance, Lütkepohl (2005) or Hamilton (1994).

The joint analysis of the time series in Figure 1.1 by means of the VAR model in Equation (1.1) might provide important evidence on their co-movement. The left-hand side panel displays the series of logarithmized energy use per capita as well as the series of logarithmized gross domestic product per capita in Australia, defining the vector $y_t = (\log(\text{energy}), \log(\text{gdp}))'$. First of all, it appears interesting to separately analyze whether the clear upward trend of the series is caused by deterministic terms, e.g. a linear trend, or if it is purely random (incorporating a drift). From visual inspection this seems not to be directly obvious. If the series move randomly, a joint consideration of both time series could clarify if they follow a joint random trend. Furthermore, the multivariate analysis enables dependencies to be uncovered and suggests causalities in one or the other direction based on their joint temporal development and dynamic relations between the series. The

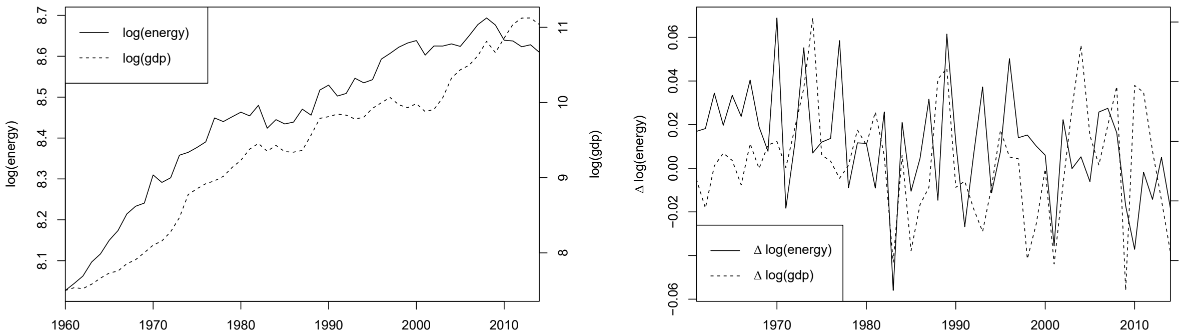


Figure 1.1: Logarithmized energy use and GDP per capita in Australia for time period 1960 to 2014 (left) and corresponding first differences (right). Note that the left and right vertical axis each provide different scales of the two variables.

series of growth rates, displayed in the right-hand side panel, feature joint positive and negative shifts more markedly. However, the direction of causation might still be unclear. Thus, the statistical analysis of the two economic time series appears crucial for gaining further insights in their economic characteristics.

In the following, we first consider a single macroeconomic time series. A series which incorporates only temporary deviations from deterministic terms provides a different starting point for economic interpretations than an ‘unpredictable’ random series. In the context of dependency, we are interested in the decision on whether the time series is stationary or exhibits persistent dependence on longer horizons. In Chapter 2 we consider univariate tests for stationarity under distinct model settings. Furthermore, panel unit root tests have been developed to overcome power deficiencies and provide a more general economic statement. While standard panel unit root tests are not robust under time-varying variances and trending data, we introduce a new test procedure which performs well in this framework. Thereby, Chapter 2 boils down to the first attached paper, which is entitled

> **Heteroskedastictiy-robust unit root testing for trending panels**
> with Helmut Herwartz and Yabibal W. Walle,
> *cege working paper Number 314, Appendix A.*

Departing from the framework of a single time series the diagnosis of dependencies between several variables provides evidence on relations in a macroeconomic system. Based on the assumption of stationarity of the series we analyze instantaneous and persistent relations between macroeconomic indices. Dependencies can, beyond the standard linear setting, be present in diverse forms. We first review the variety of dependencies in Chapter 3. Nonparametric dependence measures are developed to robustly diagnose diverse types of dependence. We compare several approaches in the first paper of this chapter,

> **Nonparametric tests for independence - a review and comparative simulation study with an application to malnutrition data in India**
> with Helmut Herwartz,
> *under review at Statistical Science, Appendix B.*

Drawing on these dependence diagnostics helps to conduct a preliminary analysis of the data. In macroeconomics, the analyst might be further interested in causalities. We analyze causalities by means of structural vector autoregressive models tracing the variables back to unanticipated independent structural shocks. We are mainly interested in the identification of the instantaneous response matrix relying on non-Gaussianity of the structural shocks. We review and compare such independence based identification procedures based on dependence measures, as studied in the first paper. Thereby, we highlight their performance by means of a simulation study and an application in

> **Independence based identification of structural shocks: Performance evaluation by means of Monte Carlo simulations and an application to the global crude oil market**
> with Helmut Herwartz,
> *Appendix C.*

For the examined approaches the assumption of at most one Gaussian (structural) shock is essential for unique identification of the instantaneous response to independent shocks. However, in the presence of multiple Gaussian shocks the non-Gaussian structural shocks can still be uniquely identified. We prove this result and, thereby, enable a more general definition of identifiability in the paper

> **Identification of independent structural shocks in the presence of multiple Gaussian components**
> *Appendix D.*

In the following, we separate the two main topics, panel unit root tests and the identification in structural VAR models, into two chapters. Therein, we introduce the theoretical basics essential for the related papers. Within the text, the papers are summarized briefly. Furthermore, we describe the individual contributions of the authors. The Appendices A to D contain the corresponding papers.

2 Persistent dependence within time series

Quantifying serial dependence within a time series can be of fundamental use when analyzing the persistence of changes in a variable. In classical time series analysis, dependence structures are usually represented by linear time-discrete autoregressive (AR) processes as defined in Equation (1.1). Standard estimation techniques are routinely applied to assess the intensity of such associations. Furthermore, test procedures allow for diagnosing dependence in the time dimension either in a linear or in nonparametric settings.[1] In addition to correlations between near epochs, the analyst might be interested in the persistence of correlations over larger time horizons. In this sense, unit root tests help to decide between a high level of correlation within the time series (unit root) or decreasing dependence for more distanced time periods (stationarity) by setting the cut-off point to unit autoregressive coefficients. In particular, the distinction between a unit root process, i.e. difference stationary, and a process fluctuating around a deterministic trend, i.e. trend stationary, might be fundamental for subsequent economic interpretations. Furthermore, many macroeconomic series can be observed for different individuals, e.g. countries. Panel unit root tests enable to arrive at the stationarity decision of the series based on panel data in a more powerful manner.

The relevance of stationarity diagnostics for macroeconomic analyses becomes apparent considering the series of energy use per capita in OECD countries (taken from Herwartz et al., 2017). The distinction between trend and difference stationarity of energy use per capita has been intensively studied within the past two decades. This might be attributed to three main reasons (Narayan and Smyth, 2007). First, knowing the direction of causality between per capita energy use and economic growth is critical for policy decisions. Since non-stationarity of the variables would have implications on testing and interpreting (causal) relations, unit root testing is routinely performed before cointegration and causality analysis between energy use and GDP per capita (displayed in Figure 1.1). Second, trend stationarity of energy use per capita is relevant for the effectiveness of energy policies. In particular, if energy consumption is a stationary process, it will return to its trend

[1] See the next chapter for more details on diverse dependence structures. Furthermore, diverse tests for independence are specialized to test for independence in time series, so called tests for serial independence. For an overview and more details see, for instance, Gooijer (2017).

after a policy shock. This implies that energy saving policies will have transitory effects only. Otherwise, if energy consumption contains a unit root, such policies will have a permanent impact. Third, stationarity of energy consumption facilitates forecasting energy demand as its past behavior offers valuable information for the prediction. However, if energy consumption is a unit root process, it does not follow a predictable path and, hence, forecasting energy demand will be more difficult than in the stationary case.[2] For further motivation and exemplary applications of unit root tests on macroeconomic variables the interested reader might consult Campbell and Perron (1991).

In the following, we first introduce the concept stationarity and how unit root tests help to decide on its presence. We consider essential basics for panel root testing in Section 2.2. Additionally, power deficiencies and related problems in dealing with non-standard data structures as varying variances and the presence of trends are discussed. Subsequently, the paper in Appendix A proposes a new heteroskedasticity-robust unit root test for trending panels.

2.1 Stationarity of time series and unit root tests

A univariate time series $\{y_t\}_{t=-\infty}^{\infty}$ comprises time-discrete realizations of one variable. Interesting characteristics of the underlying economic data are detectable in y_t to a different extent. In particular, the separation between deterministic and random components might be fundamental for further economic interpretations (as outlined for energy consumption above). For these purposes, the standard AR(1) regression model describes autoregressive structures in y_t by

$$y_t = \mu + (1-\rho)\delta t + \rho y_{t-1} + e_t, \tag{2.1}$$

where e_t is the series of error terms, δ is a trend parameter and μ contains the intercept. Note that for theoretical convenience the process is defined for all $-\infty < t < \infty$. However, the time horizon of observations might correspond to $t = 1, \dots, T$. First, the error terms e_t fulfill standard assumptions having zero mean and time-independent variance σ^2. In this setting, we determine how the standard process formulates under stationarity and non-stationarity (following Hamilton, 1994) before providing a more detailed description of unit root diagnostics. Note that the AR(1) process in (2.1) is a very stylized model description. More generally, the model would allow for lags up to order p which is not further pursued for simplicity. Further non-standard settings are postponed to later descriptions.

The model in Equation (2.1) is *covariance stationary (weak stationary)* if

$$\begin{aligned} E[y_t] &= \mu^*, \quad \text{for all } t, \\ E[(y_t-\mu)(y_{t-j}-\mu)] &= \gamma_j, \quad \text{for all } t \text{ and any } j, \end{aligned}$$

where μ^* denotes the time-independent unconditional mean and γ_j is the covariance which

[2] More detailed explanations can be found in the application part of the paper in Appendix A.

only depends on the distance $j \leq t$ between periods. The first condition implicitly excludes the linear trend term from Equation (2.1), i.e. $\delta = 0$.[3] Thus, a stationary series features identical first and second moments in all time periods. The alternative (MA(∞)) representation of the series,

$$y_t = \mu/(1-\rho) + e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \ldots, \tag{2.2}$$

gives rise to a theoretical condition for covariance stationarity. Assuming $|\rho| < 1$ in (2.2) implies absolute summability of $\sum_{j=0}^{\infty} |\rho|^j = 1/(1-|\rho|)$ and thus, a well defined stationary process y_t with unconditional mean $\mu^* = \mu/(1-\rho)$. In other words, this condition leads to decreasing dependence between the increments of $\{y_t\}_{t=1}^{T}$ for increasing time distances.[4]

Thus, both time-independent covariances and an autoregressive coefficient below one characterize stationarity. These conditions provide starting points for two formulations of non-stationarity, one of deterministic and one of random nature. First, the general model formulation in (2.1) allows for a permanent deterministic relation. If $\delta \neq 0$ and $|\rho| < 1$, the series incorporates a linear trend. Consequently, the process is not covariance stationary but has a time-dependent mean. However, filtering out the permanent deterministic shift by detrending methods leads to the remaining stationary series. Therefore, a process of this type is called *trend stationary*. Besides linear time-dependence distinct formulations of the deterministic trend (e.g. exponential or logarithmic) might be more adequate but are not further pursued here for simplicity.

Moreover, stationarity covers time series described in Equation (2.1) with autoregressive coefficients $-1 < \rho < 1$. Along these lines, non-stationarity of a random form can be caused by the presence of a *unit root*. A unit root is characterized by coefficient $\rho = 1$. This implies the following model equation:

$$y_t = \mu + y_{t-1} + e_t. \tag{2.3}$$

Most simply, for $\mu = 0$ the series corresponds to a *random walk* $y_t = \sum_{i=0}^{\infty} e_{t-i}$. For $\mu \neq 0$ the process extends to a *random walk with drift*. In difference to the stationary case the process shows no tendency for mean reversion under a unit root. By construction, the coefficient in front of the trend term equals zero in case of a unit root. As a result, the model in (2.1) rules out random walks with a trend. Nevertheless, such a model is well-formulated to distinguish between trend and difference stationarity. Alternatively, a random walk is called *integrated (of order one, I(1))* as it results from integration (the sum) over the stationary first differences of the process $\Delta y_t = y_t - y_{t-1} = e_t$.[5] After a brief illustration of these two formulations of a non-stationary process we describe adequate

[3]Beyond the first two moments, *strict stationarity* is concerned about the whole marginal and joint distribution of $(y_t, \ldots, y_{t+j})$. In the following, we concentrate on covariance stationary time series.

[4]The stationarity condition for more general models including lags up to order p corresponds to: The lag polynomial $1 - \rho_1 z - \rho_2 z^2 - \ldots - \rho_p z^p$ for autoregressive corefficients $\rho_1, \rho_2, \ldots, \rho_p$ has roots outside the unit circle $|z| > 1$ (for more detailed descriptions, see Hamilton, 1994).

[5]In general, a time series is integrated of order d if applying the difference operator d-times leads to a stationary ($I(0)$) series.

tools to diagnose stationarity in the next section.

As described above, evaluating stationarity of energy consumption is highly relevant for the further analysis of the series. Inspection of the series in Figure 2.1 underlines the question whether the process is difference stationary (i.e. contains a unit root) or trend stationary. More precisely, Figure 2.1 displays energy use per capita in Australia for time period 1960 to 2014 accompanied by its first differences $\Delta y_t = y_t - y_{t-1}$ in the right-hand side panel. Based on this, we want to decide if removing the unit root leads to a stationary process. From visual inspection the series of differences indeed appears stationary, i.e. fluctuates around its mean μ with constant variance. The associated model for the level series containing a unit root would correspond to $y_t = \mu + y_{t-1} + e_t$.

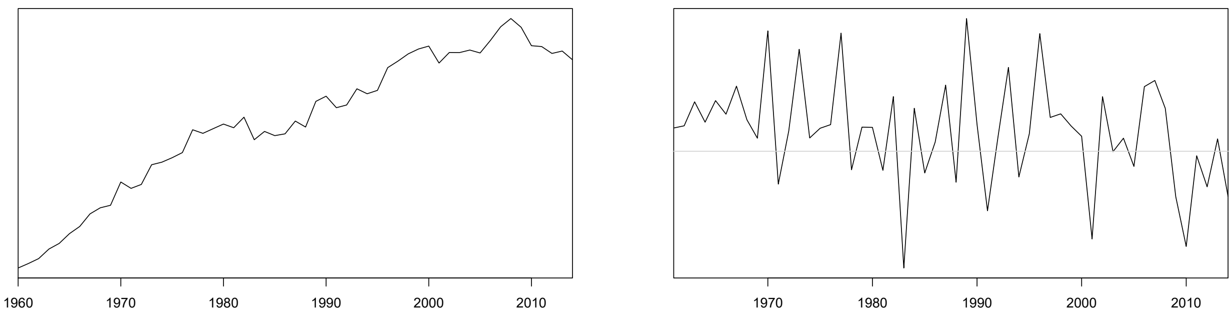
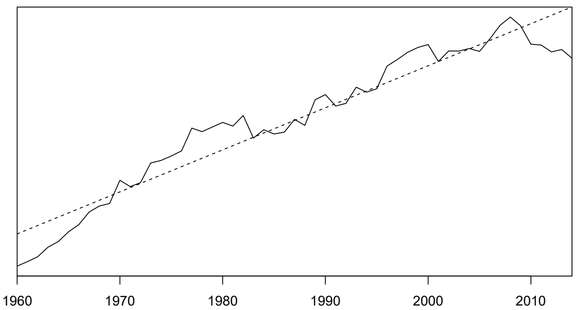


Figure 2.1: Energy use per capita in Australia for time period 1960 to 2014 (left-hand side panel) and corresponding first differences (right-hand side panel).

In addition to the original series of energy use, the left-hand side panel of Figure 2.2 displays a linear trend estimated by least squares from the regression $y_t = \mu + \delta t + e_t$. The distance between the series and the estimated linear trend, i.e. $\hat{e}_t = \hat{\mu} + \hat{\delta} t - y_t$, is displayed in the right-hand side panel. Stationarity of the series of differences would imply that the level process is trend stationary and fluctuates around the deterministic trend. It might be argued that the series in Figure 2.2 contains some structure instead of (time) independent increments. Including further autoregressive (but stationary) coefficients into the regression, i.e. a higher lag order than in the model in (2.1), could remove the remaining serial correlation. Furthermore, the trend might be expressed as a logarithmic or more generally, non-linear (but still deterministic) relation. Overall, the question whether the series is stationarity remains unanswered. Whereas stationarity has been considered only theoretically so far, an analyst might especially be interested in the practical decision in favor or against stationarity. In the following, we consider tests for unit roots enabling to formulate the characteristics of energy use in statistical terms.

2.1.1 Testing for unit roots

Diagnosing stationarity of a time series provides information in both economic and statistical sense. For the simple AR(1) model in (2.1) an appropriate test to decide between stationarity and a unit root compares the null hypothesis $H_0 : \rho = 1$ against the alternative



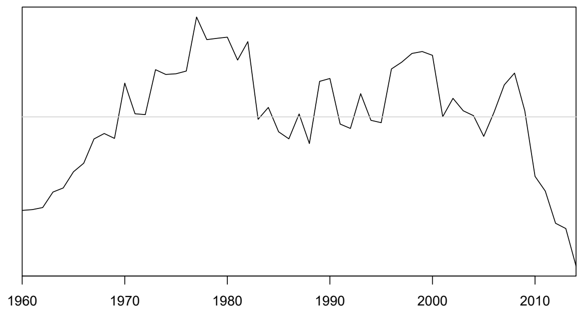


Figure 2.2: Energy use per capita in Australia for time period 1960 to 2014 with estimated linear trend (left-hand side panel) and differences from the linear trend (right-hand side panel).

$H_1 : \rho < 1$. Transforming the model in (2.1), we obtain the Dickey-Fuller regression,

$$\Delta y_t = \mu - \phi\delta t + \phi y_{t-1} + e_t, \tag{2.4}$$

where $\phi = -(1-\rho)$. Within this framework, the null hypothesis of a unit root corresponds to $H_0 : \phi = 0$ tested against the alternative hypothesis of $H_1 : \phi < 0$. Note that the scenario of an explosive series with $\phi > 0$ might also be of interest but will not be further pursued in the following (see, for instance, Chapter 10 in Fuller, 1996; Phillips et al., 2011).[6]

One might argue that the term "stationarity is often too narrowly conceived by linking it exclusively to unit root tests" (H. Rohloff, personal communication, March 2017) and that other characterizations by means of statistical or economic methodology might be of additional validity for interpretations. Nevertheless, unit root tests provide a simple and helpful assessment of stationarity providing a direct statement on the overall dependence within the time series. In the following, we describe unit root tests in the univariate case. Modifications to non-standard settings are studied thereafter.

Dickey-Fuller (DF) test

One of the most popular approaches to test for a unit root is the DF test introduced by Dickey and Fuller (1979). In the most simple case of $\mu = 0$ and $\delta = 0$ in Equation (2.4), the test statistic reads as

$$\tau_{DF} = \frac{\hat{\phi}}{\mathrm{sd}[\hat{\phi}]} \xrightarrow{d} \frac{1}{2}\frac{B(1)^2 - 1}{\sqrt{\int B(r)^2 dr}} \quad, T \to \infty, \tag{2.5}$$

[6]Extending the model in (2.1) to higher lag orders $p > 1$, we obtain the Augmented Dickey-Fuller regression, $\Delta y_t = \mu - \phi\delta t + \phi y_{t-1} + \sum_{j=1}^{p-1} \rho_j^* \Delta y_{t-j} + e_t$, where $\phi = -(1 - \rho_1 - \ldots - \rho_p)$ and $\rho_j^* = -(\rho_{j+1} + \ldots + \rho_p)$.

where $B(r)$ corresponds to a standard Brownian motion evaluated in time point $r \in [0, 1]$. The right part of Equation (2.5) describes the asymptotic *Dickey-Fuller distribution* of the test statistic for infinite sample size T. While in the context of stationarity this test is mostly applied, it shows some weaknesses in terms of size and power. On the one hand, for values of ρ close to unity (often present in economic time series) several studies argue that the DF test shows low power. On the other hand, power and size distortions can occur in small samples as the convergence rate to the asymptotic Dickey-Fuller distribution is quite slow (for more details see Maddala and Kim, 1998).

Prominent alternative tests on stationarity are the Phillips-Peron (PP) test (Phillips and Perron, 1988) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992). The PP test differs from the DF test in its control for serial correlation in the error terms (differently than the augmented DF test, see footnote 6). In difference, the KPSS test contrasts the null hypothesis of stationarity against the alternative of a unit root. Both tests are implemented in many statistical software programs in addition to the DF test. For a more detailed treatment of alternative unit root tests see, for instance, Stock (1994) and the `R` package `urca` (Pfaff et al., 2016).

Modifications under deterministic terms and varying variances

Previously, the univariate DF test for unit roots was formulated for a regression model without deterministic terms. However, macroeconomic data frequently exhibit trending behavior as seen, for instance, for energy consumption. The DF test is applicable if the model includes a constant $\mu \neq 0$ and a linear trend $\delta \neq 0$. Allowing for these specifications of an AR(1) model as in (2.1), Table 2.1 displays the corresponding model formulations under the null hypothesis of a unit root and the alternative.

	$\Delta y_t = \phi y_{t-1} + e_t$ $\mu = 0,\ \delta = 0$	$\Delta y_t = \mu + \phi y_{t-1} + e_t$, $\mu \neq 0,\ \delta = 0$	$\Delta y_t = \mu - \phi\delta t + \phi y_{t-1} + e_t$, $\mu \neq 0,\ \delta \neq 0$
$H_0 : \phi = 0$	random walk	random walk with drift	random walk with drift
$H_1 : \phi < 0$	stationary	stationary, mean $\mu/(1-\rho)$	trend stationary

Table 2.1: Characteristics of y_t under H_0 of a unit root test or H_1 under distinct model formulations of (2.1) with and without including μ and δ.

Moreover, if the deterministic terms μ and δ differ from zero, the series might need to be adjusted by demeaning and/or detrending first to ensure an adequate distribution under the null hypothesis and the alternative. As described in Stock (1994) the coefficients of the deterministic terms are most commonly estimated by regressing y_t on a constant and trend. Subsequently, these terms are subtracted to derive the demeaned and detrended series. This method for removing the trend was displayed in Figure 2.2. As we have noticed for the series of energy consumption a different detrending procedure might be more appropriate. Therefore, we will study an alternative in Section 2.2 below. For further details on the asymptotic behavior of several unit root tests under diverse scenarios we refer to Stock (1994).

Most macroeconomic time series experience occasional breaks or trending behavior in their variances. For instance, the debate on the Great Moderation underscores the fact that time-varying volatility of macroeconomic series is more of a rule rather than an exception. Allowing for heteroskedasticity in the error terms, $e_t \sim (0, \sigma_t^2)$, the standard DF test cannot be applied in a straightforward manner. In this case, the asymptotic distribution of the test statistic does not necessarily coincide with the Dickey-Fuller distribution. The main approach for a robust test statistic relies on standardizing the elements of the test statistic with respect to period-specific variances. As a consequence, the need to estimate the time-specific variance and to modify the asymptotic distribution complicates the derivation of a reasonable test procedure. Modifications of standard unit root tests allowing for time-varying variances are proposed, for instance, by Cavaliere (2004); Boswijk (2005); Cavaliere and Taylor (2007). We will return to heteroskedasticity considerations in the panel case after introducing the concept of panel unit roots tests.

2.2 Panel unit root tests

The underlying model considered in the following is a *dynamic panel model* which regresses one single target variable on its own lags while considering several individuals. This model is similar to the VAR model in Equation (1.1) with a completely different focus. Under a dynamic panel model the main aim of panel unit root tests remains to decide if the one target variable (e.g. energy use per capita) is stationary. We conclude this from evaluating the variable in multiple cross-sections. In reference to the exemplary univariate series of energy use per capita, the analysis now comprises the energy use of 23 OECD countries for 1960 to 2014 (see Figure 2.3 for an illustration).

We have briefly addressed power losses when considering univariate unit root tests. Diverse approaches for *panel unit root tests (PURTs)* have been developed to overcome these deficiencies while providing pivotal (asymptotically free from nuisance parameters, e.g., normally distributed) test statistics. The first most intuitive methods represent multivariate extensions of the DF test. They combine univariate DF tests to one new pivotal test statistic (Im et al., 2003; Maddala and Wu, 1999). However, these procedures do not account for dependence between the panels (e.g. countries). More advanced PURTs (e.g. Breitung and Das, 2005) allow for weak cross-sectional dependence but are no longer pivotal if the homoskedasticity assumption is violated. These tests are not able to capture time series with trends and varying variances as argued in Demetrescu and Hanck (2012a,b) and Herwartz et al. (2016).

In the following, we consider tests which perform properly in non-standard scenarios where the straightforward PURTs might encounter problems. For this, we first introduce the model framework and the corresponding assumptions. Inspecting the individual and joint behavior of the series in Figure 2.3 the panel model needs to allow for panel-specific

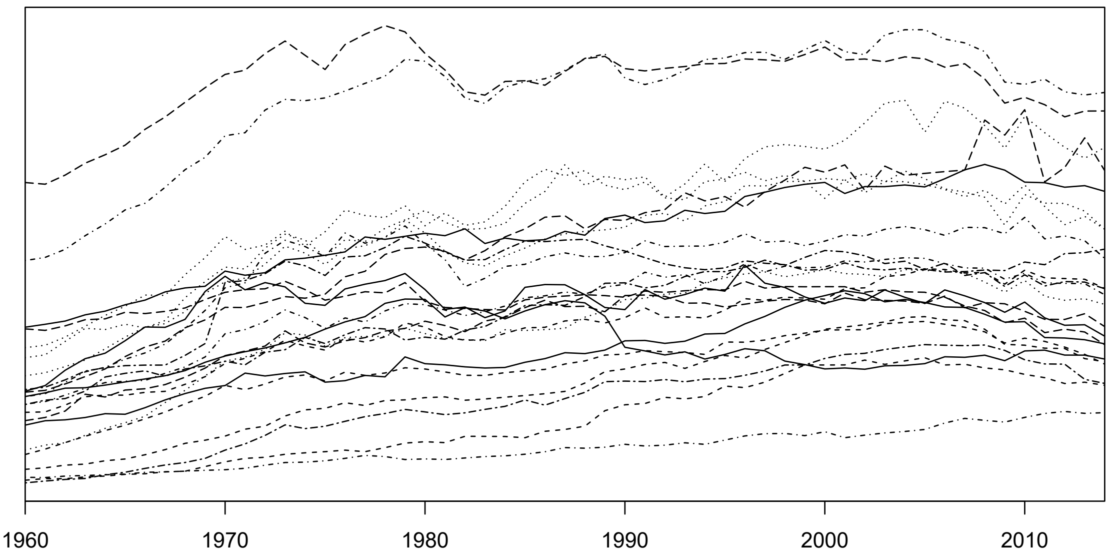


Figure 2.3: Energy use per capita in 23 OECD countries in time period 1960 to 2014

deterministic terms and cross-sectional dependence. For instance, dependence might be directly caused by relations between the sections or indirectly by common shocks in other variables. To cover this, the first order panel autoregression is specified as

$$\boldsymbol{y}_t = \boldsymbol{\mu} + (1-\rho)\boldsymbol{\delta} t + \rho \boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = 1, \ldots, T, \tag{2.6}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\boldsymbol{y}_{t-1} = (y_{1,t-1}, \ldots, y_{N,t-1})'$, $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{Nt})'$ are $N \times 1$ vectors for N distinct cross-sections. Furthermore, the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)'$ stacks individual-specific trend parameters and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$ contains individual-specific intercepts. PURTs are used to test the hypothesis $H_0 : \rho = 1$ against $H_1 : \rho < 1$ in (2.6). Along with the model specification in (2.6), we presume the following assumptions on the error terms.

Assumptions $\boldsymbol{\mathcal{A}}$.
(i) $\boldsymbol{e}_t$ is serially uncorrelated with mean vector 0 and covariance Ω_t.
(ii) Ω_t is a positive definite matrix with eigenvalues $\lambda_t^{(1)} \leq \lambda_t^{(2)} \leq \ldots \leq \lambda_t^{(N)}$ and $\lambda_t^{(N)} < \overline{c} < \infty$, $\lambda_t^{(1)} > \underline{c} > 0$ for all t and constants $\underline{c}$, $\overline{c} \in \mathbb{R}$.
(iii) $E[u_{it}^p u_{jt}^p u_{kt}^p u_{lt}^p] < \infty$ for all i, j, k, l and some $p > 1$, where $u_{\bullet t}$, $\bullet \in \{i, j, k, l\}$, denote typical elements of $\boldsymbol{u}_t = \Omega_t^{-1/2} \boldsymbol{e}_t$. Here, we set $\Omega_t^{1/2} = \Gamma_t \Lambda_t^{1/2} \Gamma_t'$ where Λ_t is a diagonal matrix of eigenvalues of Ω_t and the columns of Γ_t are the corresponding eigenvectors.

Recently, several heteroskedasticity-robust PURTs have been proposed to deal with the assumption of time-dependent covariance matrices Ω_t (Demetrescu and Hanck, 2012a,b; Herwartz et al., 2016; Westerlund, 2014). These tests also remain pivotal under cross-sectional and serial correlation as stated in (2.6) and Assumptions $\mathcal{A}$. However, they do not perform properly under detrended data. Namely, available detrending schemes introduce nuisance parameters that affect the limiting distribution of the tests under variance breaks

(Herwartz et al., 2016; Westerlund, 2014). In Section 2.2.2 we propose a panel unit root test which is robust under regression models involving heteroskedasticity and a linear trend. This PURT presupposes Assumptions $\mathcal{A}$.

Removing the trend in (2.6) by means of popular schemes such as OLS, GLS or recursive detrending renders the PURTs to depend on the drift terms in $\boldsymbol{\mu}$, and, hence, requires bias-correction terms. The newly introduced test utilizes the detrending scheme in Demetrescu and Hanck (2016). This method involves recursively detrending the lagged level variable to obtain

$$\tilde{\boldsymbol{y}}_{t-1} = \boldsymbol{y}_{t-1} + \frac{2}{t-1}\sum_{j=1}^{t-1}\boldsymbol{y}_j - \frac{6}{t(t-1)}\sum_{j=1}^{t-1} j\boldsymbol{y}_j. \tag{2.7}$$

Since $\Delta\boldsymbol{y}_t$ has non-zero mean, it has to be demeaned. One choice is to center $\Delta\boldsymbol{y}_t$ in the usual way as

$$\Delta\boldsymbol{y}_t^* = \Delta\boldsymbol{y}_t - \frac{1}{T}\sum_{t=2}^{T}\Delta\boldsymbol{y}_t. \tag{2.8}$$

We refer to the paper in Appendix A and the literature therein for more detailed descriptions of distinct detrending schemes. The modified processes in (2.7) and (2.8) provoke dependencies between increments of $\tilde{\boldsymbol{y}}_t$ in the time dimension (involving observations from past and future time periods). Consequently, for a test statistic building on these quantities standard asymptotic theory does not apply. However, the theory of near-epoch dependent processes controls for these dependencies in the time dimension. In the following, we sketch out derivations of the paper in Appendix A leading to the proof of asymptotic normality of the proposed test statistic.

2.2.1 Central limit theorem for dependent processes

If the increments of a stochastic process are not independent over time, standard asymptotic theory, as the law of large numbers or the central limit theorem, is not straightforward. Nevertheless, only allowing up to a certain level of dependence mixingales and near-epoch dependent sequences include 'asymptotically independent' increments.[7] The newly introduced test statistic can be formulated as the sum $\tau = \sum_{t=2}^{T} X_{NT,t}$ of such near-epoch dependent sequences $X_{NT,t}$ with N cross-sections and T time periods (the indices are not explicitly considered for now but become important in the actual proofs). Before arriving at the actual definition of the test statistic in the paper in Appendix A we first describe near-epoch dependence in general. For a detailed review on the theory of martingales and mixingales we refer to Davidson (1994).

The definitions of mixing and near-epoch dependent sequences are borrowed from Section 14.1 and Definition 17.1 of Davidson (1994). For a stochastic sequence $\{\boldsymbol{V}_{T,t}\}_{-\infty}^{\infty}$, possibly vector valued, on a probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{F}_{-\infty}^{t} = \sigma(\ldots, \boldsymbol{V}_{T,t-2}, \boldsymbol{V}_{T,t-1}, \boldsymbol{V}_{T,t})$

[7] For larger samples the controlled dependence within a finite number of leads and lags looses its relevance if all other periods are independent.

be the filtration generated by the sequence $\boldsymbol{V}_{T,t}$. Furthermore, the filtration $\mathcal{F}_{t+m}^{\infty} = \sigma(\boldsymbol{V}_{T,t+m}, \boldsymbol{V}_{T,t+m+1}, \boldsymbol{V}_{T,t+m+2}, \ldots)$ is similarly defined. The sequence $\boldsymbol{V}_{T,t}$ is said to be *α-mixing (or strong mixing)* if $\lim_{m\to\infty} \alpha_m = 0$ where

$$\alpha_m = \sup_t \alpha(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+m}^{\infty}) = \sup_t \sup_{A\in\mathcal{F}_{T,t+m}^{\infty}, B\in\mathcal{F}_{T,-\infty}^{t}} \Big|P(A\cap B) - P(A)P(B)\Big|. \tag{2.9}$$

Furthermore, $\boldsymbol{V}_{T,t}$ is α-mixing of size $-\varphi_0$ if $\alpha_m = \mathcal{O}(m^{-\varphi})$ for some $\varphi > \varphi_0$. Under independence $P(A\cap B) = P(A)P(B)$ so that the sequence α_m indicates to which extent the relation between the events A and B deviates from independence. For increasing m the sequence goes to zero, i.e., the two events approach independence. This formulation of 'controlled dependence' gives rise to the definition of mixingales. Consequently, limit theory for mixingales follows under certain regularity conditions (Davidson, 1994). Loosely defined, a mixingale approaches the martingale property for $m \to \infty$ while conditioning on $\mathcal{F}_{-\infty}^{t-m}$. Instead of reviewing mixingales in detail we describe the more general form of near-epoch dependent sequences.

Let $X_{NT,t} = g_t(\ldots, \boldsymbol{V}_{T,t-1}, \boldsymbol{V}_{T,t}, \boldsymbol{V}_{T,t+1}, \ldots)$ be a function of a vector of mixing processes $\boldsymbol{V}_{T,t}$. For the stochastic sequence $\{\boldsymbol{V}_{T,t}\}_{-\infty}^{+\infty}$ let $\mathcal{F}_{t-m}^{t+m} = \sigma(\boldsymbol{V}_{t-m}, \ldots, \boldsymbol{V}_{t+m})$, such that $\{\mathcal{F}_{t-m}^{t+m}\}_{m=0}^{\infty}$ is an increasing sequence of σ-fields. If, for $p > 0$, a sequence of integrable random variables $\{X_{NT,t}\}_{-\infty}^{+\infty}$ satisfies

$$||X_{NT,t} - E(X_{NT,t}|\mathcal{F}_{t-m}^{t+m})||_p \leq d_t \nu_m, \tag{2.10}$$

where $\nu_m \longrightarrow 0$, and $\{d_t\}_{-\infty}^{+\infty}$ is a sequence of positive constants, $X_{NT,t}$ will be said to be *near-epoch dependent in L_p-norm (L_p-NED)* on the mixing sequence $\{\boldsymbol{V}_t\}_{-\infty}^{+\infty}$. This means that the sequence $X_{NT,t}$ might not be mixing itself but exhibits dependence on the 'near epoch' of $\{\boldsymbol{V}_{T,t}\}$. This allows for the application of asymptotic theory similar to mixingales. In particular, a central limit theorem (CLT) for the test statistic τ holds if the following conditions of Corollary 24.7 in Davidson (1994) are fulfilled:

(a) $X_{NT,t}$ is $\mathcal{F}_{T,-\infty}^{t}$ measurable with $E\left[X_{NT,t}\right] = 0$ and $E\left[\left(\sum_{t=2}^{T} X_{NT,t}\right)^2\right] = 1$.

(b) There exists a constant array $\{c_{NT,t}\}$ such that $\sup_{T,t} ||X_{NT,t}/c_{NT,t}||_r < \infty$ for $r > 2$.

(c) $X_{NT,t}$ is L_2-NED of size -1 on $\boldsymbol{V}_{T,t}$ which is α-mixing of size $-r/(r-2)$.

(d) $\sup_T\{T(\max_{1\leq t\leq T} c_{NT,t})^2\} < \infty$.

In the paper in Appendix A we show that under Assumptions $\boldsymbol{\mathcal{A}}$ on the error terms and $N/T^2 \to 0$ the conditions (a)–(d) are fulfilled for the sequence $X_{NT,t}$ which is determined by the detrended and demeaned regression.

2.2.2 Heteroskedasticity-robust unit root testing for trending panels

joint work with Helmut Herwartz & Yabibal M. Walle, *cege discussion paper No. 314*.

In the paper in Appendix A we propose a new heteroskedasticity-robust PURT. Most importantly, the test can be applied to detrended data and its limiting distribution (under the null hypothesis) is free of nuisance parameters. The construction of the test is simple. We begin by detrending the data according to the method suggested in Demetrescu and Hanck (2016), and trace the effects of the detrending scheme on the (detrended) integrated level data. The drift term is estimated as the unconditional mean of first-differenced series. Taking account of volatility breaks, level detrending and drift estimation, we construct a test statistic that exhibits an asymptotic Gaussian distribution under the panel unit root null hypothesis. To prove asymptotic normality we rely on central limit theory for near-epoch dependent processes as discussed, e.g., in Davidson (1994). Simulation results show that the proposed test works well in finite samples, and has satisfactory power which is comparable with the power of the tests in Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2012a) under homoskedasticity.

As an empirical illustration, we examine whether energy use per capita is trend or difference stationary. Using data from 23 OECD economies over the period 1960–2014, we find that energy use per capita is generally integrated of order one. However, results from unit root testing for rolling fixed-length time spans show that the series could be characterized as trend stationary for forty-years windows that start between 1963 and 1968.

My contributions to the paper are the following:

- I was mainly responsible for the mathematical proofs of the asymptotic results. With respect to the present test statistic this involves the handling of central limit theory of the so-called near-epoch dependent processes.
- Related to asymptotic normality of the test statistic, several preliminary lemmas contain consistency results and results on stochastic orders. I was in charge for stating and proving these lemmas.
- I wrote and structured theoretical passages of the paper as well as the whole appendix.

The work of H. Herwartz and Y. Walle include the development of a suitable test statistic as well as preliminary work in the field of panel unit root tests. Y. Walle was mainly responsible for the simulation study and the application to the series of energy consumption. Furthermore, H. Herwartz supported the derivation of the proof by means of proof-reading and mathematical discussions.

3 Dependence between time series

In the previous chapter, (panel) unit root tests have proven to be a useful tool to decide on the persistence of dependence structures within a time series. In the following, we leave the dimension of one single series and assume a priori that the considered series does not exhibit a unit root (or is already transformed to be stationary). Switching to the connection axis between distinct quantities of interest enables to analyze either linear or various alternative types of dependence. In a first step, we investigate this diversity of dependence structures. Additionally, nonparametric dependence measures are presented as robust tools to diagnose the significance of diverse relations between random variables.

Determining (in)dependence between economic quantities might not be straightforward. The non-experimental nature of, e.g. macroeconomic data, complicates clear statements about relations between variables. For instance, dependence over a third (omitted) variable might provoke two variables to be ‘indirectly’ dependent. Moreover, the endogeneity of most macroeconomic variables makes a precise analysis of the structures within such a system interesting. Correlations between one target and several explanatory variables over a certain time horizon are traditionally studied by means of standard regression models. Without prior knowledge about causalities between the quantities, the application of a vector autoregressive (VAR) model, as in Equation (1.1), leaves their relationships more agnostic. This enables to conduct causality analyses between macroeconomic indices in several ways. For instance, structural vector autoregressive (SVAR) models serve to identify driving factors for changes in macroeconomic variables. Thereby, the changes can be traced back to independent (or uncorrelated) structural shocks. Here, nonparametric dependence measures provide an important means to decide if the non-unique uncorrelated structural shocks are independent and, specifically, to determine those which are least dependent.

In the following section, we briefly address the diversity of dependence structures potentially present in a set of random variables. Subsequently, Section 3.2 and the associated paper in Appendix B review important approaches to diagnose these structures nonparametrically. We highlight their merits and drawbacks by comparing them within a simulation study. Furthermore, after a brief introduction on causality analysis we assess causality in sets of macroeconomic variables by means of SVAR models. Thereby, we separate sources by means of nonparametric dependence measures in Section 3.3.1 and the respective paper in Appendix C. Lastly, the paper in 3.3.2 and Appendix D refers to a further extension of the independence based identification procedures.

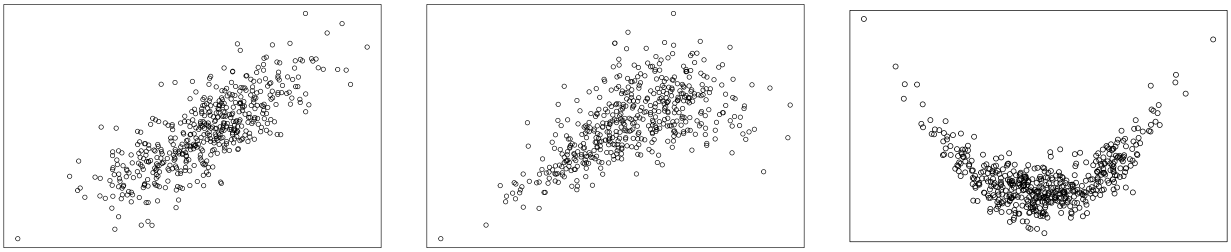


Figure 3.1: Bivariate standard normal distribution with $\rho = 0.8$ (left), normally distributed variables with Clayton copula with parameter $\theta = 1.5$ (middle) and the functional relationship $x_2 = x_1^2 + \varepsilon$ for $x_1 \sim \mathcal{N}(0, 0.5)$ and $\varepsilon \sim \mathcal{N}(0, 0.2)$ (right). *Figure taken from Herwartz and Maxand (2017).*

3.1 Dependence structures

In a multivariate (macroeconomic) system a central objective of analyses is the relationship between the included quantities. In this sense, the following definition is essential to proceed.

Definition 1. *Two random variables $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$ with associated distribution functions F_{x_1}, F_{x_2} and joint distribution function F are independent if and only if*

$$F(x_1, x_2) = F_{x_1}(x_1) F_{x_2}(x_2). \tag{3.1}$$

Definition 1 implicitly determines *dependence* as the complement. Despite describing independence by means of distribution functions F, the (separate and joint) densities and characteristic functions provide a basis for equivalent formulations. Mistakenly, the similar concept of *correlation* is often treated as a synonym for dependence. Two random variables x_1 and x_2 are correlated if their covariance differs from zero. Under the standard assumption of a multivariate normal distribution these two terms coincide as the normal distribution is fully described by its first two moments. However, in distinct distributional settings, uncorrelated variables can still exhibit diverse forms of dependence.

Figure 3.1 displays three representative relationships between x_1 and x_2. The left-hand side panel displays linear dependence within a bivariate normal distribution. Furthermore, dependence in the lower tail of the distributions characterizes the second structure. Thirdly, a functional nonlinear and nonmonotone association relates the variables in the right-hand side panel of Figure 3.1. Increasing the number of considered variables also raises the number and complexity of possible relations connecting them. Along these lines, a multivariate set $\{x_1, \ldots, x_p\}$ of univariate random variables $x_1, \ldots, x_p \in \mathbb{R}$ provides room for diverse forms of dependence. In the paper in Appendix B, we describe bivariate, groupwise and mutual dependence of $p > 2$ random variables in more detail. An analyst aims to decide if a set of observed variables exhibits specific types of or dependence at all. Nonparametric dependence measures can prove useful for these purposes.

3.2 Nonparametric dependence measures

Pearson's correlation coefficient (see, e.g. Pearson, 1920) might be the most noted dependence (or, more precisely, correlation) measure. For samples of univariate random variables $\{x_{1,1}, \ldots, x_{1,n}\}$ and $\{x_{2,1}, \ldots, x_{2,n}\}$ this coefficient accounts for linear bivariate structures by

$$\rho_{x_1,x_2} = \frac{\sum_{i=1}^{n}(x_{1,i} - \overline{x}_1)(x_{2,i} - \overline{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{1,i} - \overline{x}_1)}\sqrt{\sum_{i=1}^{n}(x_{2,i} - \overline{x}_2)}} \tag{3.2}$$

where $\overline{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1,i}$ and $\overline{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{2,i}$ are the sample means. The correlation coefficients of the samples displayed in Figure 3.1 correspond to $\rho_1 = 0.81$, $\rho_2 = 0.63$ and $\rho_3 = 0.01$ (from left to right). Even if the correlation coefficient for the data in the right-hand side panel is small, the data displays a clear structure. Apparently, the correlation coefficient does not properly represent this form of nonlinear and nonmonotone dependence. Additionally, the tail dependence in the second panel of Figure 3.1 provokes a lower dependence coefficient than the standard normal distributed data on the left. However, it is not clear if the smaller coefficient is caused by not perceiving the nonlinearity of the dependence structure or by an overall smaller degree of dependence.

With regard to the diversity of dependence structures the informational content of standard linear dependence measures, such as Pearson's correlation coefficient, might be limited. As an alternative, nonparametric dependence measures aim at detecting diverse forms of dependence while keeping prior assumptions on the distribution and the dependence structure at a minimum. Scanning the literature on nonparametric tests of independence, we encounter a 'zoo' of alternative approaches. The underlying null hypothesis of the tests for bivariate dependence between random variables x_1 and x_2 formulates as

$$H_0 : F(x_1, x_2) = F_{x_1}(x_1)F_{x_2}(x_2) \quad \text{vs.} \quad H_1 : F(x_1, x_2) \neq F_{x_1}(x_1)F_{x_2}(x_2). \tag{3.3}$$

Classical rank correlation methods include Kendall's tau (Kendall, 1938) and Spearman's rho (Spearman, 1904) to test nonparametrically for independence in bivariate settings. Extending the bivariate measures to diagnose dependence in multivariate sets of random variables is, however, more sophisticated. In this regard, independence diagnostics might be classified into four distinct categories according to their theoretical background. Besides traditional approaches, copula, spatial rank and kernel-based methods have been developed more recently to test nonparametrically for independence in a multivariate framework. The next section as well as the paper in Appendix B provide a comprehensive review and comparison of the wide range of nonparametric independence tests.

3.2.1 Nonparametric tests for independence – a review and comparative simulation study with an application to malnutrition data in India

joint work with Helmut Herwartz, under review at *Statistical Science*.

The paper in Appendix B reviews available approaches to test nonparametrically for independence in bivariate as well as multivariate settings. It offers a compact overview on the diversity of existing tests concentrating on those which are ready-to-use (i.e., implemented in `R` packages). Even if nonparametric tests of the null hypothesis of independence aim at performing adequately irrespective of the underlying distribution, they rely on certain (test specific) regularity assumptions. To identify sources of differences in performance, we first review the theoretical background of the test procedures. Moreover, we consider their performance under specific marginal distributions and dependence structures by means of a simulation study. We distinguish diverse nonmonotone and nonlinear dependence structures generated by copulas and based on functional associations. Additionally, for specific applications, e.g., economic data, modifications of these structures might be of interest. The variety of nonparametric independence tests shows distinguished performances in terms of empirical size and power in various small sample settings. As a result, we help on deciding which test to use with regard to representative underlying distributional settings. Based on distinguished test outcomes in small samples, we detect nonlinear dependence structures between childhood malnutrition indices and possible determinants in an empirical application for India.

My contributions to the paper are the following:

- I reviewed the literature on alternative independence tests. This work includes the selection of the most interesting candidates for a simulation study and the careful shortening of theoretical results from the original papers.

- I was responsible for the simulation study including the choice of the most informative simulation settings and evaluation of test performance measures.

- I conducted the investigation of Indian malnutrition data by means of the selected independence tests.

- I wrote the draft for the whole paper and was mainly responsible for the alterations made to the manuscript.

The work of H. Herwartz includes advisory support in conducting the simulation study. Furthermore, extensive proof-reading work as well as the (re)formulation of several passages has substantially improved the readability of the paper.

3.3 Causality analysis by means of structural VAR models

Dependence measures help to diagnose and quantify the dependence between variables as described in the last section. Especially in economics, the analyst might, beyond mere dependence, be interested in causal relations. First of all, a proper definition of the term causality is needed to proceed without misconception with respect to the identification and interpretation of causal structures. However, a universal definition would require to diverge into philosophy and is beyond the scope of this thesis. Therefore, we apply a stylized characterization of causality. A significant relation between two macroeconomic variables indicates dependence in the first place. Furthermore, we add the direction of causation as sort of a target value defining directed dependence, i.e. *causality*. At the same time, bilateral relations are still possible.

In macroeconomic analysis causality considerations help to understand the driving forces of economic variables (for an overview, see, Moneta et al., 2011; Hoover, 2001). In particular, causes can be rooted in shocks associated with macroeconomic indices. This helps to decide if manipulation of the variable of interest (the gross domestic product is a typical example) is possible by turning the correct screws of the economic machinery. Monetary policy and its influences on the economy is a much discussed and prominent example for such an adjustment screw (see, for instance, Uhlig, 2005). For the detection of causalities between macroeconomic indices it might be noteworthy that the data is non-experimental and observed. Hence, we can hardly control if all causing variables are included in the regression model which implies that direct causes and confounding might not be observable. However, the feasibility of finding causal relations in an abstracted model is never certain but most often presumed for simplicity.

Prior economic beliefs of causality directions are implemented in a standard regression model. Since Sims (1980), vector autoregressive (VAR) models as formulated in (1.1) are commonly used to study causal relations in macroeconomics allowing for bilateral feedback relations. Based on a VAR model, Granger (1980) approaches causality in an intuitive way. One variable is said to be *Granger-causal* for another variable if inclusion of the causing variable to the information set (everything known at time t) significantly changes the forecast of the caused variable in $t+1$. However, this concept of causality is based on the observed errors of the model instead of assessing structural components present in terms of instantaneous correlation. Therefore, actual causality might not be assessed as argued by Hoover (2001). *Structural VAR* (SVAR) models provide an alternative framework for the assessment of causal relations. Deriving a two dimensional SVAR model from two simple one-directional regressions in the following, assesses SVAR models in an intuitive manner. For illustration purposes, a model without any lag structure or constant terms allows to concentrate on the determination of causalities. Along these lines, the main interest lies on the identification of instantaneous relations ('causality'). The incorporation of dynamic structures is rather straightforward and will be considered afterwards. In the framework of a simple linear regression model the two variants of contrary causality are

$$y_{1t} = \alpha_{12} y_{2t} + \sigma_1 \varepsilon_{1t}, \tag{3.4}$$

$$y_{2t} = \alpha_{21} y_{1t} + \sigma_2 \varepsilon_{2t}, \quad t = 1, \ldots, T \tag{3.5}$$

where $\varepsilon_{kt} \sim (0,1), k = 1,2$, are uncorrelated unit error terms with corresponding standard deviations σ_1 and σ_2. Utilizing either Equation (3.4) or (3.5) to estimate the relation between the two variables implicitly implements the direction of correlation by not allowing for a bilateral relation. In a more general formulation, the vector form of the two models allows for simultaneous estimation of the effects corresponding to

$$\begin{aligned}
\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 0 & \alpha_{12} \\ \alpha_{21} & 0 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} + \begin{pmatrix} \sigma_1 \varepsilon_{1t} \\ \sigma_2 \varepsilon_{2t} \end{pmatrix} \\
\begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} \sigma_1 \varepsilon_{1t} \\ \sigma_2 \varepsilon_{2t} \end{pmatrix} \\
\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \\
&:= \mathsf{B} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.
\end{aligned} \tag{3.6}$$

The matrix in the second row is invertible if its determinant is different from zero, i.e. $1 - \alpha_{21}\alpha_{12} \neq 0$ which implies $\alpha_{21} \neq 1/\alpha_{12}$. Under this condition the structural model is called fundamental (or causal). In the following, we assume fundamentalness of the model but briefly revert to nonfundamentalness after introducing autoregressive and deterministic terms into the model. The matrix B in (3.6) represents the instantaneous relation between (and within) the two variables. For instance, the interpretation of the upper right entry B_{12} of the matrix would be as follows: A unit shock to the first equation causes an instantaneous increase by B_{12} in the level of the second variable. Next, we extend the model with lags from preceding time periods and deterministic terms to capture the autoregressive behavior and allow for a trend and intercept term. This reveals the connection between the model in (3.6) and the one in Equation (1.1). We derive the corresponding K-dimensional VAR model of order p as

$$\begin{aligned}
y_t &= c_t + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \\
&= c_t + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \mathsf{B}\varepsilon_t, \\
\Leftrightarrow \quad A(L) y_t &= c_t + \mathsf{B}\varepsilon_t, \qquad t = 1, \ldots, T,
\end{aligned} \tag{3.7}$$

with vector valued deterministic terms c_t and $A(L) = I - A_1 L - \ldots - A_p L^p$, where I denotes the $K \times K$ identity matrix. As mentioned below Equation (3.6), we assume the model to be fundamental. A fundamental model with autoregressive terms is characterized by an invertible matrix $A(L)$, i.e. $\det(A(z)) \neq 0$ for all $|z| \leq 1$. For instance, Alessi et al. (2008) and Hansen and Sargent (1980) describe that nonfundamentalness might arise due to model misspecification and omitted variables. However, in the present stylized framework we assume to obtain a proper model and therefore, do not further pursue this issue.

The stochastic model components in (3.7) are commonly characterized from two perspectives: Firstly, *reduced form residuals* u_t correspond to error terms with zero mean $E(u_t) = 0$ and covariance matrix $\Sigma_u = \mathsf{B}\mathsf{B}'$. Secondly, *structural shocks* $\varepsilon_t = \mathsf{B}^{-1}u_t$ are uncorrelated unit shocks with $E(\varepsilon_t) = 0$ and $\Sigma_\varepsilon = \mathsf{B}^{-1}\Sigma_u\mathsf{B}^{-1} = I_K$. Without any further restrictions, the factor B of the covariance matrix Σ_u is not unique such that different choices of B are observationally equivalent. Along these lines, estimating the two directions of correlation simultaneously and leaving the causal direction agnostic leads to identification problems. For interpretation of the impact of structural shocks the matrix B, however, has to be identified properly.

Identification in SVAR models

The literature on SVAR models incorporates diverse approaches to solve (or at least reduce) the identification problem assuming either statistical or economic properties of the structural shocks (for a textbook treatment of SVARs see Kilian and Lütkepohl, 2017). Comonly, changes in macroeconomic variables are traced back to uncorrelated structural shocks $\varepsilon_t = \mathsf{B}^{-1}u_t$. Uncorrelated shocks are derived by all decompositions B of the covariance matrix, i.e. the matrix B and the associated structural shocks ε_t are not unique. Traditional identification procedures of structural VAR models aim at reducing the non-uniqueness of the identification. They apply economically motivated restrictions to derive an identified set of matrices. More precisely, the whole set of possible covariance decompositions is restricted to those in line with prior economic beliefs. For instance, Sims (1980) and Blanchard and Quah (1989) implement instantaneous zero and long-run restrictions while Faust (1998); Canova and Nicolo (2002); Uhlig (2005) restrict the entries of the decomposition B to display predetermined signs.

One could argue that analyzing the impact of uncorrelated structural shocks omits potential dependencies in distributions different from Gaussianity (as mentioned in Section 3.1).[1] Additionally, deviations from a stringent normality assumption provide a powerful tool for unique identification of the structural shocks. Along these lines, several recent approaches have exploited further statistical characteristics of the underlying data structure and the distribution of u_t. Deviating from pure normality, for instance, Lanne and Lütkepohl (2010) derive at identification of the model assuming a mixture of two Gaussian distributions. Similar to Rigobon (2003) and Lewbel (2010), Lanne and Lütkepohl (2008) propose an identification scheme that builds upon time-heterogeneous covariance estimators. Herwartz and Lütkepohl (2014) show how such heteroskedasticity-based identification approaches can be combined with external information derived from economic

[1] (S)VAR models formulate dependencies between time series in a linear framework. Covering the diversity of potential dependence structures by means of alternative models might be more involved and way beyond the scope of this thesis. For instance, nonlinear model formulations are described in Lütkepohl (2005).

theory. In the following, we aim at the complete separation of structural shocks to interpret the impact of a certain shock properly. Only independent signals enable to clearly trace back changes in variables to their sources. If the structural shocks, i.e. $\varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon)$, are Gaussian, no correlation directly implies independence. However, the immaculate assumption of a Gaussian distribution might be too restrictive, or simply not correct, for the data at hand.

Against this background, we are mainly concerned with independence based identification. Source separation under non-Gaussianity leads to the field of *independent component analysis (ICA)*. Originally developed for signal processing, ICA techniques help to determine the matrix B leading to independent shocks ε_t. A comprehensive overview on ICA approaches can be found in Hyvärinen et al. (2001). The development of these techniques mostly rests upon the central result of Comon (1994) who shows that the matrix B can be uniquely identified if at most one shock $\varepsilon_{kt}, k = 1, \ldots, K$, is Gaussian. As a consequence, ICA algorithms are useful to identify B assuming that ε_t contains at most one Gaussian component. However, in the paper in Appendix D we show that identification of non-Gaussian components is also possible if ε_t contains multiple Gaussian components. First, the paper in Appendix C describes several identification procedures basing on ICA in more detail. Thereby, we find the way back to nonparametric dependence measures which enable to determine least dependent shocks without restricting or excluding any distribution or type of dependence. Before turning to the exact formulation and comparison of the procedures we briefly review the tools used for interpretation of the identified matrices.

Techniques for SVAR analysis

After proper identification of B we can proceed by calculating the dynamic and instantaneous impact of structural shocks on a macroeconomic system. Diverse tools serve to analyze the relations by highlighting specific characteristics (see, for instance, Lütkepohl, 2011). *Impulse response functions* describe the impact of unit shocks with respect to a certain response delay (delay of zero gives the instantaneous impact). Future effects of an economic shock are observed in the reaction of the variables included in the model. For the model formulation in (3.7) the response matrices are derived from

$$\begin{aligned} A(L)y_t &= c_t + \mathsf{B}\varepsilon_t \\ y_t &= A(L)^{-1}c_t + A(L)^{-1}\mathsf{B}\varepsilon_t \\ &= \mu_t + \Phi(L)\mathsf{B}\varepsilon_t = \mu_t + \sum_{i=0}^{\infty} \Phi_i \mathsf{B}\varepsilon_{t-i} = \mu_t + \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} \end{aligned}$$

where μ_t represents the unconditional mean of the series. Matrix $\Theta_i := \Phi_i\mathsf{B}$ is the moving average matrix composed of the dynamic relation Φ_i and the instantaneous correlation matrix B. In particular, $\Theta_0 = \mathsf{B}$.

As an extension, *forecast error variance decompositions (FEVD)* give the relative importance of each shock for the variations of the considered variables. For the multivariate series y_t the corresponding h-step ahead forecast error is defined as $y_{t+h} - y_{t|t}(h) = \Theta_0 \varepsilon_{t+h} + \ldots + \Theta_h \varepsilon_{t+1}$. The forecast error variance of the kth variable is $\sigma_k^2(h) = \sum_{j=0}^{h-1}(\Theta_{k1,j}^2 + \ldots + \Theta_{kK,j}^2)$. Noting that $\Sigma_\varepsilon = I_K$, the relative contribution of shock j is consequently defined as

$$FEVD_{kj}(h) = (\Theta_{kj,0}^2 + \ldots + \Theta_{kj,h-1}^2)/\sigma_k^2(h). \tag{3.8}$$

Further inference about the contribution of structural shocks to the variable of interest can be drawn from *historical decompositions.* The contribution of shock j to variable k in time period t is

$$y_{jt}^{(k)} = \sum_{i=0}^{t-1} \Theta_{kj,i} \varepsilon_{k,t-i} + \alpha_{j1}^{(t)} y_0 + \ldots + \alpha_{jp}^{(t)} y_{-p+1}, \tag{3.9}$$

where $\alpha_{ji}^{(t)}$ is the jth row of $A_i^{(t)}$ and $[A_1^{(t)}, \ldots, A_p^{(t)}]$ consists of the first K rows of the companion form matrix with exponent t, $\boldsymbol{A}^t$ (see Lütkepohl, 2005, for more details).

3.3.1 Independence based identification of structural shocks: Performance evaluation by means of Monte Carlo simulations and an application to the global crude oil market

joint work with Helmut Herwartz.

To identify the contemporaneous linkages among reduced form disturbances to an interplay of orthogonal structural shocks of unit variance, the SVAR analysis has to rely on additional (often external and not data-based) information. We compare the performance of three alternative independence based identification procedures and identification by means of sign restrictions under distinct distributional settings and sample sizes. We resume the widely applied identification procedure based on sign restrictions. Furthermore, three approaches based on non-normality of structural shocks are considered. The first procedure has been advocated in Lanne et al. (2017) and is based on maximum likelihood (ML) estimation assuming, for instance, t-distributed structural shocks. Relaxing the strict distributional assumptions required for ML estimation, two further identification strategies allow an interpretation as Hodges Lehman (HL) estimation of the structural model (Hodges and Lehmann, 2006). Principles of HL estimation motivate the detection of least dependent structural shocks by the minimization of two alternative nonparametric dependence criteria, namely the so-called distance covariance of Bakirov et al. (2006) and the Cramér-von Mises distance of Genest and Rémillard (2004). While the former has already been employed in the context of independent component analysis (Matteson and Tsay, 2013), the latter has been suggested for point estimation of cyclic SVARs by Herwartz (2015). Within a simulation study, we confirm a bias induced by stylized sign restrictions and find considerable differences between parametric and nonparametric identification schemes that exploit the supposed independence of structural shocks. In an application to the global crude oil market independence based identification performs comparable with techniques of former studies without the need to set up strong economic or distributional assumptions a-priori.

My contributions to the paper are the following:

- I was responsible for the simulation study including the choice of the identification procedures, their implementation in `R` and the evaluation by means of MSE and the sign pattern.
- I was in charge of the application of the chosen procedures to the oil data set.
- I wrote the draft for the whole paper and was responsible for alterations made to the manuscript.

The work of H. Herwartz includes instructing assistance with respect to the simulation study and application. Moreover, proof-reading and the (re)formulation of certain passages improved the paper substantially.

3.3.2 Identification of independent structural shocks in the presence of multiple Gaussian components

Several recently developed identification techniques for structural VAR models base on the assumption of non-Gaussianity. So-called independence based identification provides unique structural shocks (up to scaling and ordering) under the assumption of at most one Gaussian component. While non-Gaussianity of certain interesting shocks, e.g., a monetary policy shock, appears rather natural, not all macroeconomic shocks in the system might show this clear difference from Gaussianity. We generalize identifiability by noting that even in the presence of multiple Gaussian shocks the non-Gaussian ones are still unique. Consequently, independence based identification allows to uniquely determine the (non-Gaussian) shock of interest irrespective of the distribution of the remaining system. In an illustrative macroeconomic model the identified structural shocks confirm the results of previous studies on the early millennium slowdown. Furthermore, estimation based on extended time horizons enables empirical evidence for the whole model being fully identified under the non-Gaussianity assumption.

I am the single author of this paper and responsible for all the written text, the theoretical statements and empirical results.

4 Conclusions and outlook

In the following, we briefly summarize the main contributions made in the two chapters of this dissertation. Additionally, we deduce consecutive research directions and give a long run outlook.

Panel unit root tests

Unit root tests provide a straightforward assessment of the persistence of changes in time series. In particular, heteroskedasticity-robust panel unit root testing is an important research topic in panel data econometrics. The paper in Appendix A proposes a new panel unit root test (PURT) which performs well for trending heteroskedastic panels. Among rival approaches this test is unique in obtaining a pivotal test statistic under general (and typical) features of macroeconomic panel data. Noticing this important progress, the author's current research on panel unit root tests has already been proceeding. A STATA implementation of recent heteroskedastictiy-robust PURTs includes the one proposed in Appendix A and enables straightforward accessibility of the tests. The STATA package is available on request and the corresponding manual is already submitted (joint with H. Herwartz, F. Raters, Y. Walle).

The detrending procedure described in Section 2.2 induces a rather involved variance estimation. In a subsequent paper (in preparation) we alternatively propose the GLS detrending scheme suggested in Demetrescu and Hanck (2016) which relies on nonparametric estimation of the varying cross-section specific variances as described in Boswijk (2005). Asymptotic properties will be proven and merits and drawbacks of the two alternative detrending procedure highlighted. Moreover, the assumption of weak cross-sectional dependence is necessary for the PURT in Appendix A to perform properly. As might be argued, cross-sectional data often involves stronger forms of dependence as, for instance, caused by factor structures. Consequently, the assumptions on the covariance matrix in Section 2.2 might be too restrictive in many macroeconomic applications (see Andrews, 2005, for further discussions on this issue). Further modifications of recent test procedures might lead to robustness under stronger forms of dependence between heteroskedastic panels with trend. As a first step, evaluating the relevance of the assumptions on the period-specific covariance matrix might provide valuable information on how to adjust the test statistics accordingly.

Nonparametric tests for independence

We have seen that nonparametric dependence measures provide robust tools to analyze dependencies in various sets of variables. We describe performance properties of representative nonparametric approaches in the paper in Appendix B by comparing diverse tests with respect to a variety of different settings. While we consider the tests in model settings of two and three random variables, the dependence structures can become increasingly complicated in larger sets of variables. Furthermore, the literature on nonparametric independence tests is growing, and already provides refinements of the methods discussed in this work. Thus, it appears to be a fruitful avenue for future research to characterize merits and risks of most recent dependence diagnostics under diverse distributional settings and higher dimensionality by means of simulation studies.

Identification in structural VAR models

Several recent papers argue that independent component analysis (ICA) proves useful for the identification of structural vector autoregressive (SVAR) models. In the paper in Appendix C, we specifically highlight the performance of three alternative independence based identification procedures illustrated by means of macroeconomic applications. This provides a convenient starting point for future work on these identification procedures. First, the incorporation of more recent approaches relying on non-Gaussian structural shocks such as Gouriéroux et al. (2017) or Capasso and Moneta (2016) could instructively extend the simulation study. Additionally, utilizing the studied approaches to assess much discussed economic questions underlines their merits (and drawbacks) and enables economic interpretations without strong a-priori restrictions. For instance, investigating the monetary policy asset price nexus by applying the distance covariance for identification provides economic reasonable instantaneous and dynamic responses in a five and six dimensional model. The corresponding paper is in preparation (joint with H. Herwartz and H. Rohloff). Moreover, a universally usable implementation in `R` (under construction) will enable the flexible application of independence based identification.

Implementing additional properties of ICA techniques (see, for instance, Hyvärinen et al., 2001) to modify existing identification approaches promises to reduce remaining estimation and interpretation ambiguities within the SVAR analysis. In this sense, future research can pursue two major objectives. The first one appears more specific: While independence based identification allows for the estimation of responses to independent structural shocks uniquely up to scaling, the economic interpretation still remains challenging. More precisely, identifying independent shocks does not yet assign a reasonable economic label and meaning. This also causes difficulties in the calculation of confidence intervals. However, measuring the accuracy of the identified structural matrices plays an important role to detect economic causalities, and in particular, to decide on the significance of an impact. The consequent research might target the theoretical and simulation based assessment of appropriate resampling techniques in combination with independence based identification. Besides improved resampling techniques, alternative methods to label the structural shocks might be evaluated by comparing their benefit for reasonable economic interpretations.

More comprehensively, the methodological enhancement of the considered identification procedures can base on two concepts utilizing either further economic or statistical data characteristics. In former works, statistical identification of non-Gaussian structural shocks has been implemented straightforwardly to SVAR analysis. However, certain difficulties remain. For instance, in the paper in Appendix D we show that samples with multiple Gaussian disturbances allow to uniquely identify the non-Gaussian shocks whereas the Gaussian ones can not be distinguished from each other. Therefore, it might be beneficial to link independence based identification to common economic intuitions to rectify these interpretative deficiencies. Implemented in sign restrictions or zero restrictions, well-founded economic theory can help to separate two Gaussian structural shocks, or in the same way, shocks which are close to normality. Besides economic properties, the performance of ICA algorithms promises to be improved by exploiting additional data characteristics of the structural shocks. Hyvärinen (2013) provides an overview on recent advances in ICA techniques which might help to develop modified identification approaches. Utilizing, for instance, volatility shifts or time-dependencies of the error terms might help to decrease the estimation bias of the structural matrix, especially in small samples, high dimensions or data-rich environments.

Bibliography

Alessi, L., Barigozzi, M., Capasso, M., 2008. A review of nonfundamentalness and identification in structural var models. Working Paper 922, European Central Bank.

Andrews, D., 2005. Cross-section regression with common shocks. Econometrica 73 (5), 1551–1585.

Bakirov, N. K., Rizzo, M. L., Székely, G. J., 2006. A multivariate nonparametric test of independence. J. Multivariate Anal. 97 (8), 1742–1756.

Blanchard, O. J., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. American Economic Review 79 (4), 655–73.

Boswijk, H. P., 2005. Adaptive testing for a unit root with nonstationary volatility. UvA-Econometrics Discussion Paper 2005/07.

Breitung, J., Das, S., 2005. Panel unit root tests under cross sectional dependence. Statistica Neerlandica 59 (4), 414–433.

Campbell, J., Perron, P., 1991. Pitfalls and opportunities: What macroeconomists should know about unit roots. In: NBER Macroeconomics Annual 1991, Volume 6. National Bureau of Economic Research, Inc, pp. 141–220.

Canova, F., Nicolo, G. D., 2002. Monetary disturbances matter for business fluctuations in the g-7. Journal of Monetary Economics 49 (6), 1131–1159.

Capasso, M., Moneta, A., 2016. Macroeconomic responses to an independent monetary policy shock: a (more) agnostic identification procedure. Lem papers series, Sant'Anna School of Advanced Studies, Pisa, Italy.

Cavaliere, G., 2004. Unit root tests under time varying variances. Econometric Reviews 23 (4), 259–292.

Cavaliere, G., Taylor, A. M. R., 2007. Testing for unit roots in time series with non-stationary volatility. Journal of Econometrics 140 (2), 919–947.

Comon, P., 1994. Independent component analysis, A new concept? Signal Processing 36 (3), 287–314.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

Demetrescu, M., Hanck, C., 2012a. A simple nonstationary-volatility robust panel unit root test. Economics Letters 117 (2), 10–13.

Demetrescu, M., Hanck, C., 2012b. Unit root testing in heteroscedastic panels using the Cauchy estimator. Journal of Business and Economic Statistics 30 (2), 256–264.

Demetrescu, M., Hanck, C., 2016. Robust inference for near-unit root processes with time-varying error variances. Econometric Reviews 35 (5), 751–781.

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

Faust, J., 1998. The robustness of identified var conclusions about money. Carnegie-Rochester Conference Series on Public Policy 49 (1), 207–244.

Fuller, W. A., 1996. Introduction to statistical time series, 2nd Edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, a Wiley-Interscience Publication.

Genest, C., Rémillard, B., 2004. Tests of independence and randomness based on the empirical copula process. Test 13 (2), 335–370.

Gooijer, J. G. D., 2017. Elements of Nonlinear Time Series Analysis and Forecasting. Springer International Publishing Switzerland 2017, Ch. Tests for serial independence, pp. 257–314.

Gouriéroux, C., Monfort, A., Renne, J.-P., 2017. Statistical inference for independent component analysis: Application to structural var models. Journal of Econometrics 196, 111–126.

Granger, C. W. J., 1980. Testing for causality: a personal viewpoint. J. Econom. Dynamics Control 2 (4), 329–352.

Hamilton, J. D., 1994. Time series analysis. Princeton University Press, Princeton, NJ.

Hansen, L. P., Sargent, T. J., 1980. Formulating and estimating dynamic linear rational expectations models. J. Econom. Dynamics Control 2 (1), 7–46.

Herwartz, H., 2015. Structural var modelling with independent innovations - an analysis of macroeconomic dynamics in the euro area based on a novel identification scheme, mimeo.

Herwartz, H., Lütkepohl, H., 2014. Structural vector autoregressions with Markov switching: combining conventional with statistical identification of shocks. J. Econometrics 183 (1), 104–116.

Herwartz, H., Maxand, S., 2017. Nonparametric tests for independence - a review and comparative simulation study with an application to malnutrition data in india, mimeo.

Herwartz, H., Maxand, S., Walle, Y., 2017. Heteroskedasticity-robust unit root testing for trending panels. cege discussion paper 314.

Herwartz, H., Siedenburg, F., 2008. Homogenous panel unit root tests under cross sectional dependence: Finite sample modifications and the wild bootstrap. Computational Statistics and Data Analysis 53 (1), 137–150.

Herwartz, H., Siedenburg, F., Walle, Y. M., 2016. Heteroskedasticity robust panel unit root testing under variance breaks in pooled regressions. Econometric Reviews 35 (5), 727–750.

Hodges, J., Lehmann, E., 2006. Hodges-lehmann estimators. In: Encyclopedia of Statistical Sciences.

Hoover, K., 2001. Causality in Macroeconomics. Cambridge University Press.

Hyvärinen, A., 2013. Independent component analysis: recent advances. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 371(1984).

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons.

Im, K. S., Pesaran, H. M., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115 (1), 53–74.

Kendall, M. G., 1938. A new measure of rank correlation. Biometrika 30 (1/2), 81–93.

Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press, forthcoming.

Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics 54 (1-3), 159–178.

Lanne, M., Lütkepohl, H., 2008. Identifying monetary policy shocks via changes in volatility. Journal of Money, Credit and Banking 40 (09), 1131–1149.

Lanne, M., Lütkepohl, H., 2010. Structural vector autoregressions with nonnormal residuals. J. Bus. Econom. Statist. 28 (1), 159–168.

Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. J. Econometrics 196 (2), 288–304.

Lewbel, A., 2010. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. Journal of Business & Economic Statistics 30 (1), 67–80.

Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer-Verlag, Berlin.

Lütkepohl, H., 2011. Vector autoregressive models. Eui working paper eco 2011/30, European University Institute, Florence.

Maddala, G. S., Kim, I.-M., 1998. Unit Roots, Cointegration, and Structural Change. Cambridge University Press.

Maddala, G. S., Wu, S., 1999. A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and Statistics 61 (0), 631–52.

Matteson, D. S., Tsay, R. S., 2013. Independent component analysis via distance covariance. pre-printHttp://arxiv.org/abs/1306.4911.

Moneta, A., Chlaß, N., Entner, D., Hoyer, P., 2011. Causal search in structural vector autoregressive models. In: JMLR: Workshop and Conference Proceedings 12. pp. 95–118.

Narayan, P. K., Smyth, R., 2007. Are shocks to energy consumption permanent or temporary? Evidence from 182 countries. Energy Policy 35 (1), 333–341.

Pearson, K., 1920. Notes on the history of correlation. Biometrika 13 (1).

Pfaff, B., Zivot, E., Stigler, M., 2016. urca: Unit Root and Cointegration Tests for Time Series Data. R package version 1.3.

Phillips, P., Wu, Y., Yu, J., 2011. Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? International Economic Review 52 (1), 201–226.

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

Rigobon, R., 2003. Identification through heteroskedasticity. The Review of Economics and Statistics 85, 777–792.

Sims, C. A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.

Spearman, C., 1904. The proof and measurement of association between two things. American Journal of Psychology 15, 72–101.

Stock, J. H., 1994. Unit roots, structural breaks and trends. In: Handbook of econometrics, Vol. IV. Vol. 2 of Handbooks in Econom. North-Holland, Amsterdam, pp. 2739–2841.

Uhlig, H., 2005. What are the effects of monetary policy on output? results from an agnostic identification procedure. Journal of Monetary Economics 52 (2), 381–419.

Westerlund, J., 2014. Heteroscedasticity robust panel unit root tests. Journal of Business & Economic Statistics 32 (1), 112–135.

A Heteroskedasticity-robust unit root testing for trending panels

Heteroskedasticity-robust unit root testing for trending panels

Helmut Herwartz*, Simone Maxand and Yabibal M. Walle

Georg-August-University Göttingen, Germany.

Abstract Standard panel unit root tests (PURTs) are not robust to breaks in innovation variances. Consequently, recent papers have proposed PURTs that are pivotal in the presence of volatility shifts. The applicability of these tests, however, has been restricted to cases where the data contains only an intercept, and not a linear trend. This paper proposes a new heteroskedasticity-robust PURT that works well for trending data. Under the null hypothesis, the test statistic has a limiting Gaussian distribution. Simulation results reveal that the test tends to be conservative but shows remarkable power in finite samples.

JEL Classification: C23, C12, Q40.

Keywords: Panel unit root tests, nonstationary volatility, cross-sectional dependence, near epoch dependence, energy use per capita.

*Address correspondence to: Helmut Herwartz, Chair of Econometrics, Georg-August-University Göttingen, Humboldtallee 3, D-37073 Göttingen, Germany, Telephone: +49 551 39 27350, fax: +49 551 397279, e-mail: *hherwartz@uni-goettingen.de*.

1 Introduction

Most macroeconomic time series experience occasional breaks or trending behavior in their unconditional variances. For instance, Sensier and van Dijk (2004) document that, during the period 1959–1999, about 80% of 214 U.S. macroeconomic time series they studied displayed breaks in their unconditional volatility. It is also well-known that volatilities of several macroeconomic series were significantly lower during the period 1984–2007 than in earlier decades, a phenomenon called the 'Great Moderation' (see, for instance, Stock and Watson, 2003). However, business cycle volatilities rose again during the recent global economic and financial crises. Whether the 'great recession' marks the end of the Great Moderation or was just a short interruption within an ongoing Great Moderation is still debated.[1] In any case, the debate—or even the very notion of Great Moderation for that matter—underscores the fact that time-varying volatility of macroeconomic series is more of a rule rather than an exception.

The potential consequences of variance shifts on univariate unit root tests have been investigated by, among others, Hamori and Tokihisa (1997), Kim et al. (2002), Cavaliere (2004), and Cavaliere and Taylor (2007). These studies find that the (augmented) Dickey-Fuller (Dickey and Fuller, 1979) tests have seriously distorted empirical sizes—and, hence, provide deceptive inference—if volatility varies over time. The same problem carries over to panel unit root tests (PURTs), as shown in Demetrescu and Hanck (2012a,b) and Herwartz et al. (2016). In particular, widely applied PURTs such as those suggested in Levin et al. (2002) and Breitung and Das (2005) are no longer pivotal if the homoskedasticity assumption is violated (Herwartz et al., 2016).

To deal with the above problem, a few heteroskedasticity-robust PURTs have been proposed recently. In consecutive papers, Demetrescu and Hanck (2012a,b) suggest PURTs that are built on the so-called Cauchy estimator. As the sign function of Cauchy instrumenting reduces the lagged level series to -1 and 1—irrespective of the underlying time varying volatility—these tests are argued to be robust to heteroskedasticity. Herwartz et al. (2016) show that the non-Cauchy version of the test in Demetrescu and Hanck (2012a), which was initially proposed in Herwartz and Siedenburg (2008), is robust to volatility shifts. Another heteroskedasticity-robust PURT has been suggested by Westerlund (2014). This test utilizes the information contained in group-specific

[1] See, for instance, Gadea-Rivas et al. (2014) for a concise survey on this debate.

variances.

While these heteroskedasticity-robust tests also remain pivotal under a fairly general form of cross-sectional and serial correlation, they, however, do not work for detrended data. Namely, available detrending schemes introduce nuisance parameters that affect the limiting distribution of the tests under variance breaks (Herwartz et al., 2016; Westerlund, 2014). This problem significantly limits the applicability of the tests as many macroeconomic time series exhibit trending behavior. In fact, Westerlund (2015, p. 454) states that

> "*...for many economic time series, a linear trend, rather than a constant, might be considered appropriate as the default specification, ... This is certainly true for series such as GDP, industrial production, money supply and consumer or commodity prices, where trending behavior is evident.*"

In this paper, we propose a new heteroskedasticity-robust PURT. Most importantly, the test can be applied to detrended data and its limiting distribution (under the null hypothesis) is free of nuisance parameters. The construction of the test is simple. We begin by detrending the data according to the method suggested in Demetrescu and Hanck (2014), and trace the effects of the detrending scheme on the (detrended) integrated level data. The drift term is estimated as the unconditional mean of first-differenced series. Taking account of volatility breaks, level detrending and drift estimation, we construct a test statistic that exhibits an asymptotic Gaussian distribution under the panel unit root null hypothesis. To prove asymptotic normality we rely on central limit theory for near-epoch dependent processes as discussed, e.g., in Davidson (1994). Simulation results show that the proposed test works well in finite samples, and has satisfactory power which is comparable with the power of the tests in Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2012a) under homoskedasticity.

As an empirical illustration, we examine whether energy use per capita is trend or difference stationary. Using data from 23 OECD economies over the period 1960–2014, we find that energy use per capita is generally integrated of order one. However, results from unit root testing for rolling fixed-length time spans show that the series could be characterized as trend stationary for forty-years windows that start between 1963 and 1968.

Section 2 sketches the panel unit root testing problem and describes two of the existing heteroskedasticity-robust PURTs. Section 3 discusses ways of handling serial correlation and deterministic terms. Section 4 introduces the proposed test statistic and states its asymptotic distribution. The finite sample performance of the new test is evaluated by means of a Monte Carlo study documented in Section 5. As an empirical illustration, the stationarity of energy use per capita is examined in Section 6. Section 7 concludes. Proofs of the asymptotic results are provided in the Appendix.

2 Homogeneous panel unit root testing

In this section we first describe the panel unit root testing problem and formalize cross-sectional dependence and heteroskedasticity. Next, we present the White-type heteroskedasticity-robust PURTs suggested in Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2012a).

2.1 The first order panel autoregression

A first order panel autoregression under nonstationary volatility and a linear trend can be specified as

$$\boldsymbol{y}_t = \boldsymbol{\mu} + (1-\rho)\boldsymbol{\delta}t + \rho\boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = 1, ..., T, \tag{1}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\boldsymbol{y}_{t-1} = (y_{1,t-1}, \ldots, y_{N,t-1})'$, $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{Nt})'$ are $N \times 1$ vectors, and $\boldsymbol{e}_t$ is heterogeneously distributed with mean zero and covariance Ω_t. Furthermore, the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)'$ stacks panel-specific trend parameters, and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$ contains panel-specific intercepts. The specification in (1) formalizes an empirically relevant panel unit root testing problem of distinguishing between a random walk with drift on the one hand and a trend stationary process on the other hand (Pesaran, 2007). PURTs are used to test the hypothesis $H_0 : \rho = 1$ against $H_1 : \rho < 1$ in (1).

To formalize cross-sectional dependence and heteroskedasticity, we adopt the following assumptions about the vector of error terms $\boldsymbol{e}_t$ as in Herwartz et al. (2016) with strengthened moment conditions:

Assumptions $\boldsymbol{\mathcal{A}}$.

(i) $\boldsymbol{e}_t$ *is serially uncorrelated with mean* 0 *and covariance* Ω_t.

(ii) Ω_t *is a positive definite matrix with eigenvalues* $\lambda_t^{(1)} \leq \lambda_t^{(2)} \leq \ldots \leq \lambda_t^{(N)}$ *and* $\lambda_t^{(N)} < \overline{c} < \infty$, $\lambda_t^{(1)} > \underline{c} > 0$ *for all* t.

(iii) $E[u_{it}^p u_{jt}^p u_{kt}^p u_{lt}^p] < \infty$ *for all* i, j, k, l *and* $p = 1, 2$, *where* $u_{\bullet t}$, $\bullet \in \{i, j, k, l\}$ *denote typical elements of* $\boldsymbol{u}_t = \Omega_t^{-1/2}\boldsymbol{e}_t$. *Here we set* $\Omega_t^{1/2} = \Gamma_t \Lambda_t^{1/2} \Gamma_t'$, *where* Λ_t *is a diagonal matrix of eigenvalues of* Ω_t *and the columns of* Γ_t *are the corresponding eigenvectors.*

$\mathcal{A}(i)$ restricts the error terms to be serially uncorrelated. Ways of handling higher order serial correlation will be described later. The assumption that the fourth order moments of e_{it} (or u_{it} by implication of $\mathcal{A}(ii)$) should be finite ($\mathcal{A}(iii)$ for $p = 1$) is standard in the (panel) unit root literature. The stronger assumption of finiteness of moments up to order eight ($p = 2$) will allow to apply asymptotic theory for near-epoch dependent processes. While $\mathcal{A}(ii)$ captures so-called weak forms of cross-sectional dependence such as spatial panel models (for more details on spatial panel models see, e.g., Anselin, 2013) and seemingly unrelated regressions, it rules out strong forms of cross-sectional dependence that might be traced back to the presence of common factors. Since $\text{tr}(\Omega_t) = \sum_{i=1}^{N} \lambda_t^{(i)}$, $\mathcal{A}(ii)$ covers both discrete covariance breaks as well as smoothly trending variances.

2.2 Heteroskedasticity-robust tests

2.2.1 The White-type test

Herwartz and Siedenburg (2008) propose a PURT based on a White-type covariance estimator. Setting $\boldsymbol{\mu} = \boldsymbol{\delta} = 0$ in (1), the test statistic is given by

$$t_{HS} = \frac{\sum_{t=1}^{T} \boldsymbol{y}_{t-1}' \Delta \boldsymbol{y}_t}{\sqrt{\sum_{t=1}^{T} \boldsymbol{y}_{t-1}' \hat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}_t' \boldsymbol{y}_{t-1}}} \xrightarrow{d} N(0,1), \quad \hat{\boldsymbol{e}}_t = \Delta \boldsymbol{y}_t = \boldsymbol{e}_t. \tag{2}$$

Originally, t_{HS} was proposed as an alternative to the test in Breitung and Das (2005) for finite samples where the cross-sectional dimension is relatively large in comparison with the time series dimension. Recently, Herwartz et al. (2016) show that time-varying volatility does not affect the pivotalness of t_{HS}.

2.2.2 The White-type Cauchy test

Demetrescu and Hanck (2012a) suggest a heteroskedasticity-robust PURT based on the

'Cauchy' estimator which instruments the lagged level by its sign. This statistic reads as

$$t_{DH} = \frac{\sum_{t=1}^{T} sgn(\boldsymbol{y}_{t-1})'\Delta \boldsymbol{y}_t}{\sqrt{\sum_{t=1}^{T} sgn(\boldsymbol{y}_{t-1})'\hat{\boldsymbol{e}}_t\hat{\boldsymbol{e}}_t' sgn(\boldsymbol{y}_{t-1})}} \xrightarrow{d} N(0,1), \tag{3}$$

where $sgn(\cdot)$ denotes the sign function.

Two further heteroskedasticity-robust PURTs that we are aware of are those proposed in Demetrescu and Hanck (2012b) and Westerlund (2014). A common limitation of all these PURTs, however, is that in the presence of linear trends (i.e., $\boldsymbol{\delta} \neq 0$ in (1)), applying standard detrending schemes does not retain the pivotalness of the tests if the data exhibit variance breaks.

3 Deterministic terms and serial correlation

In this section, we discuss how serial correlation and deterministic terms are handled in panel unit root testing under variance breaks.

3.1 Short-run dynamics

To eliminate short-run serial correlation from the data, prewhitening is an important procedure which leaves the limiting distribution of the tests unaffected (Breitung and Das, 2005). This procedure requires estimating individual-specific autoregressions of the first differences under H_0, i.e.,

$$\Delta y_{it} = \sum_{j=1}^{p_i} b_{ij}\Delta y_{i,t-j} + e_{it}. \tag{4}$$

Prewhitened data is then obtained as

$$\widehat{y}_{it} = y_{it} - \widehat{b}_{i1}y_{i,t-1} - \ldots - \widehat{b}_{ip_i}y_{i,t-p_i}, \tag{5}$$

and

$$\widehat{\Delta y}_{it} = \Delta y_{it} - \widehat{b}_{i1}\Delta y_{i,t-1} - \ldots - \widehat{b}_{ip_i}\Delta y_{i,t-p_i}. \tag{6}$$

Any consistent lag-length selection criterion can be applied to decide upon the lag orders p_i. In cases where both short-run dynamics and deterministic patterns are present in the data, prewhitening should precede detrending. The prewhitening regression should include an intercept term if the model features linear time trends under the alternative hypothesis.

3.2 Deterministic terms

Removing the trend in (1) by means of popular schemes such as OLS, GLS or recursive detrending renders the PURTs to depend on the drift terms in $\boldsymbol{\mu}$, and, hence, requires bias-correction terms. Moreover, the bias-correction becomes highly complicated with the presence of variance breaks. The detrending procedures in Breitung and Das (2005) and Demetrescu and Hanck (2014) do not require bias adjustment terms as long as the homoskedasticity assumption is maintained. With time-varying volatility, however, both detrending methods affect the pivotalness of PURTs, including t_{HS} and t_{DH}. As the test we are proposing utilizes the detrending scheme in Demetrescu and Hanck (2014), we briefly outline it here. This method involves recursively detrending the lagged level variable to obtain

$$\tilde{\boldsymbol{y}}_{t-1} = \boldsymbol{y}_{t-1} + \frac{2}{t-1}\sum_{j=1}^{t-1}\boldsymbol{y_j} - \frac{6}{t(t-1)}\sum_{j=1}^{t-1} j\boldsymbol{y}_j. \tag{7}$$

Since $\Delta\boldsymbol{y_t}$ has non-zero mean, it has to be demeaned. One choice is to center $\Delta\boldsymbol{y_t}$ in the usual way as

$$\Delta\boldsymbol{y}_t^* = \Delta\boldsymbol{y_t} - \frac{1}{T}\sum_{t=2}^{T}\Delta\boldsymbol{y}_t, \tag{8}$$

where T in the denominator replaces $T-1$ for notational convenience. Demetrescu and Hanck (2014) show that, under homoskedasticity, $\Delta\boldsymbol{y_t}$ could also be centered by means of forward demeaning instead of (8). In the presence of heteroskedasticity, both full sample centering and forward demeaning affect the pivotalness of even the heteroskedasticity-robust tests t_{HS} and t_{DH} and, hence, invoke marked size distortions (see Demetrescu and Hanck (2014) for rigorous arguments on this issue). As forward demeaning additionally leads to relatively large power losses in comparison with full sample centering, the test proposed in this work relies on full sample demeaning.

4 Panel unit root test for trending series with time-varying volatility

The heteroskedasticity-robust test we propose builds upon the White-type test given in (2) and the detrending scheme described by (7) and (8). Instead of providing the test statistic in a compact form, we first consider a modified version of the numerator of t_{HS}

in (2). With (7) and (8) the summands of the numerator of this modification can be rewritten as

$$\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^* = \tilde{\boldsymbol{y}}_{t-1}'\hat{\boldsymbol{e}}_t = \sum_{i=1}^{t-1}\left(a_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \frac{1}{T}a_{i,t-1}\sum_{k=2}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\right), \tag{9}$$

where

$$\Delta\boldsymbol{y}_t^* = \hat{\boldsymbol{e}}_t = \boldsymbol{e}_t - \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{e}_t, \tag{10}$$

and finite weighting coefficients $a_{i,t-1}$ read as

$$a_{i,t-1} = 1 + \frac{2}{t-1}(t-i) - 3\left(1 - \frac{(i-1)i}{(t-1)t}\right). \tag{11}$$

Derived from data detrended according to (7) and (8), the expression in (9) has a non-zero expectation in the absence of homoskedasticity under the null hypothesis of a panel unit root. The theoretical version of the new test statistic, henceforth denoted by τ, can be seen as a modification of t_{HS} with adjustments for the non-zero mean in the numerator, and corresponding changes for the variance (in the denominator). Specifically, the test statistic with theoretical moments is given by

$$\tau = \frac{\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}\left(\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^* - \nu_t\right)}{\sqrt{\frac{1}{NT}\left(E\left[\sum_{t=2}^{T}\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*\right]^2 - \left(\sum_{t=2}^{T}\nu_t\right)^2\right)}}, \tag{12}$$

where $\nu_t = E[\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*]$.

Unlike in Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2014), where the White-type covariance estimator is applied, the more complicated form of $\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*$ invokes the following representation of the variance of the numerator in (12):

$$s_{NT}^2 := \frac{1}{NT}\left(E\left[\sum_{t=2}^{T}\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*\right]^2 - \left(\sum_{t=2}^{T}\nu_t\right)^2\right) = \zeta_1 - \zeta_2 + \zeta_3 + \zeta_4 + \zeta_5 - \frac{1}{NT}\left(\sum_{t=2}^{T}\nu_t\right)^2. \tag{13}$$

The expansion of the expectation in (13) yields components $\zeta_1, \ldots, \zeta_5$ which can be shown to correspond to the following quantities

$$
\begin{aligned}
\zeta_1 &= \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}\left(\mathrm{tr}(\Omega_i\Omega_j)+\mathrm{tr}(\Omega_i)\mathrm{tr}(\Omega_j)\right)\\
\zeta_2 &= \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=i+1}^{T}\tilde{a}_{i,s-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i\Omega_s)\\
\zeta_3 &= \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\tilde{a}^2_{i,t-1}\mathrm{tr}(\Omega_i\Omega_t) \qquad (14)\\
\zeta_4 &= \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\bar{a}^2_{i,t-1}\left(E\left[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2\right]+\sum_{j=1,j\neq i,t}^{T}\mathrm{tr}(\Omega_i\Omega_j)\right)\\
\zeta_5 &= \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(E\left[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2\right]+\sum_{j=1,j\neq i,t,s}^{T}\mathrm{tr}(\Omega_i\Omega_j)\right),
\end{aligned}
$$

where $\tilde{a}_{i,t-1} = \left(1-\frac{1}{T}\right)a_{i,t-1}$ and $\bar{a}_{i,t-1} = \frac{1}{T}a_{i,t-1}$ with coefficients $a_{i,t-1}$ defined in (11). Similarly,

$$\nu_t = E[\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*] = -\sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i). \qquad (15)$$

The new test is then the empirical version of τ in (12), i.e.,

$$\hat{\tau} = \frac{\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}\left(\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^* - \hat{\nu}_t\right)}{\hat{s}_{NT}}, \qquad (16)$$

where estimators of ν_t and the variance components are based on the estimation of the traces of the covariance matrices Ω_i. More precisely, we replace $\mathrm{tr}(\Omega_i)$ by $\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i$, $\mathrm{tr}(\Omega_i\Omega_j)$ by $\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j$ and $E\left[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2\right]$ by $(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2$ where $\hat{\boldsymbol{e}}_i$ is a vector of centered residuals (first differences) as defined in (10). Detailed representations of $\hat{\nu}_t$ and $\hat{s}^2_{NT}$ are given in the Appendix. The following proposition states the asymptotic normality of the statistic in (16).

Proposition 1. *Under assumptions $\mathcal{A}$ the test statistic in (16) is asymptotic normally distributed, i.e., for $N, T \to \infty$ with $N/T^2 \to 0$*

$$\hat{\tau} \xrightarrow{d} \mathcal{N}(0,1). \qquad (17)$$

The proof of Proposition 1 is based on a central limit theorem for near-epoch dependent sequences and is given in the Appendix. As it will turn out, the additional requirement of $N/T^2 \to 0$ is necessary for $\hat{\tau}$ to fulfill the conditions of the central limit theorem, as well as for applying the test to prewhitened data (Herwartz et al., 2016).

5 Monte Carlo study

5.1 The simulation design

To evaluate the finite sample properties of the proposed test $\hat{\tau}$, we consider the following DGPs taken from Pesaran (2007):

$$\text{DGP1:} \quad \boldsymbol{y}_t = \boldsymbol{\mu} + (\boldsymbol{j} - \boldsymbol{\rho}) \odot \boldsymbol{\beta} t + \boldsymbol{\rho} \odot \boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = -50, \ldots, T, \tag{18}$$

$$\text{DGP2:} \quad \boldsymbol{y}_t = \boldsymbol{\mu} + (\boldsymbol{j} - \boldsymbol{\rho}) \odot \boldsymbol{\beta} t + \boldsymbol{\rho} \odot \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t = \boldsymbol{b} \odot \boldsymbol{\epsilon}_{t-1} + \boldsymbol{e}_t, \tag{19}$$

where bold entries indicate vectors of dimension $N \times 1$, $\boldsymbol{j}$ is a vector of ones and $\odot$ denotes the Hadamard product. The DGP1 formalizes AR(1) models with serially uncorrelated innovations while DGP2 introduces AR(1) disturbances. Both DGPs formalize a panel random walk with drift under the null hypothesis, and a panel of trend stationary processes with individual effects under the alternative. Empirical size is obtained by setting $\boldsymbol{\rho} = \boldsymbol{j}$ and power is simulated as $\boldsymbol{\rho} = 0.9\boldsymbol{j}$.[2] Individual effects, trend parameters as well as serial correlation of innovations are modeled as in Pesaran (2007): $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$, $\mu_i \sim \text{iid}\, U(0, 0.02)$ and $\boldsymbol{b} = (b_1, \ldots, b_N)'$, $b_i \sim \text{iid}\, U(0.2, 0.4)$.

To separate the issue of cross-sectional correlation from variance breaks, we employ the decomposition

$$\Omega_t = \Phi_t^{1/2} \Psi \Phi_t^{1/2},$$

where $\Phi_t = diag(\sigma_{1t}^2, \ldots, \sigma_{Nt}^2)$ and Ψ is a (time invariant) correlation matrix characterizing Ω_t. Cross-sectional independence is obtained by setting Ψ to an identity matrix of order N. We generate a weak form of cross-sectional correlation by means of the spatial autoregressive (SAR) error structure used in Herwartz and Siedenburg (2008). Specifically, we take Ψ_{SAR} that is implied by the SAR model

$$\boldsymbol{e}_t = (I_N - \Theta W)^{-1} \boldsymbol{\xi}_t, \quad \text{with} \quad \Theta = 0.8 \quad \text{and} \quad \boldsymbol{\xi}_t \sim \text{iid}\, N(\boldsymbol{0}, I_N),$$

where W is the so-called spatial weights matrix. In this particular case, W is a row normalized symmetric contiguity matrix of the '*g ahead and g behind*' structure, with

[2] Results for DGPs with heterogeneous autogregressive coefficients under the alternative hypothesis, i.e., $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_N), \rho_i \sim \text{iid}\, U(0.85, 0.95)$, are qualitatively identical and available upon request. Moreover, recent papers, e.g., Homm and Breitung (2012), also consider power against explosive alternatives ($\rho > 1$). Using a right-sided testing, the proposed test $\hat{\tau}$ is powerful against the alternative that $\boldsymbol{\rho} = 1.03\boldsymbol{j}$, even for $T = 25$. The corresponding simulation results are available upon request.

$g = 1$ (see, e.g., Kelejian and Prucha, 1999). The resulting covariance matrix of $\boldsymbol{e}_t$ is given by $\Omega_{SAR} = ((I_N - \Theta W)'(I_N - \Theta W))^{-1}$, and Ψ_{SAR} is the correlation matrix implied by Ω_{SAR}.

Cross-section specific volatility shifts are generated as

$$\sigma_{it}^2 = \begin{cases} \sigma_{i1}^2, & \text{if } t < \lfloor \gamma_i T \rfloor, \ (0 < \gamma_i < 1) \\ \sigma_{i2}^2, & \text{otherwise,} \end{cases}$$

where γ_i refers to the time a variance break occurs and $\lfloor \gamma_i T \rfloor$ denotes the integer part of $\gamma_i T$. In the homoskedastic case, $\sigma_{i1} = \sigma_{i2} = 1$. We introduce heteroskedasticity by changing the post-break variance to $\sigma_{i2} = 1/3$, for a negative variance break, and to $\sigma_{i2} = 3$, for a positive one. Regarding the timing of the variance breaks, we consider scenarios of homogeneously early ($\gamma_i = 0.2$) or late ($\gamma_i = 0.8$) variance breaks for all panel units.[3] Data are generated for all combinations of $N \in [50, 100, 250]$ and $T \in [25, 50, 100, 250]$. To mitigate the potential impacts of initial values on our analysis, we generate and discard 50 presample observations.

5.2 Simulation results

In the following we discuss simulation results on the finite sample performance of the proposed test statistic $\hat{\tau}$ in comparison with two of the existing heteroskedasticity-robust tests (t_{HS} and t_{DH}). For the new test, we also document results for its theoretical counterpart τ determined from the true covariance matrices Ω_t (see (12)). Presenting simulation results for both $\hat{\tau}$ and τ is meant to highlight finite sample performance of $\hat{\tau}$ that can be traced back to the use of moment estimators.

5.2.1 Cross-sectionally independent panels

Simulation results for data generated according to DGP1 for cross-sectionally independent panels are documented in Table 1. Results in the upper panel of this table show that, under homoskedasticity, the recursive detrending scheme in Demetrescu and Hanck (2014) leaves the pivotalness of heteroskedsticity-robust tests unaffected. With respect to rejection frequencies under the alternative hypothesis, it can be seen that using estimated

[3] Main findings of the simulation exercise remain qualitatively unaffected by consideration of randomly distinct break moments $\gamma_i \sim \text{iid}\, U(0.1, 0.9)$. These results are available upon request.

Table 1: Empirical rejection frequencies, cross-sectionally independent panels

		5%								10%							
		size				power				size				power			
N	T	τ	$\hat{\tau}$	HS	DH	τ	$\hat{\tau}$	HS	DH	τ	$\hat{\tau}$	HS	DH	τ	$\hat{\tau}$	HS	DH
Constant variance (HOM)																	
50	25	5.6	4.6	4.8	4.5	40.5	32.0	31.1	19.4	11.1	10.0	9.8	9.9	55.2	49.9	48.3	32.3
50	50	4.6	5.5	4.8	4.3	98.9	98.6	97.4	77.2	10.4	11.1	9.7	9.0	99.6	99.7	99.4	87.7
50	100	4.6	4.3	3.5	4.0	100.0	100.0	100.0	100.0	9.7	9.7	8.4	8.6	100.0	100.0	100.0	100.0
50	250	3.8	4.2	3.6	4.0	100.0	100.0	100.0	100.0	9.0	9.9	8.6	8.6	100.0	100.0	100.0	100.0
100	25	5.4	3.4	4.3	4.7	63.8	50.8	49.6	28.9	11.4	8.2	9.5	9.5	76.1	68.4	67.8	45.6
100	50	5.2	5.5	4.7	4.7	100.0	100.0	100.0	96.2	11.0	10.8	9.7	9.9	100.0	100.0	100.0	98.8
100	100	4.5	4.9	3.7	4.4	100.0	100.0	100.0	100.0	9.4	10.2	8.8	8.7	100.0	100.0	100.0	100.0
100	250	4.9	4.6	3.7	4.4	100.0	100.0	100.0	100.0	10.3	10.0	8.4	9.0	100.0	100.0	100.0	100.0
250	25	5.4	2.0	4.5	4.8	92.7	81.0	83.2	54.2	10.8	5.9	9.5	9.7	96.1	92.1	93.4	70.1
250	50	4.9	4.1	3.8	4.2	100.0	100.0	100.0	100.0	10.4	9.5	9.0	9.2	100.0	100.0	100.0	100.0
250	100	4.7	4.7	3.6	4.1	100.0	100.0	100.0	100.0	9.6	9.5	8.0	8.7	100.0	100.0	100.0	100.0
250	250	4.6	5.0	3.9	4.0	100.0	100.0	100.0	100.0	10.1	10.3	8.4	9.0	100.0	100.0	100.0	100.0
Early negative variance shift (NEG)																	
50	25	4.9	1.4	0.0	0.0	9.6	4.3	0.0	0.0	10.2	4.4	0.0	0.0	17.3	12.1	0.0	0.0
50	50	5.1	3.0	0.0	0.0	48.8	48.2	0.0	0.1	10.7	7.9	0.0	0.0	63.8	66.2	0.1	0.6
50	100	4.7	4.0	0.0	0.0	99.9	99.9	37.8	36.1	10.0	8.9	0.0	0.0	100.0	100.0	54.4	51.4
50	250	4.6	4.5	0.0	0.0	100.0	100.0	100.0	100.0	9.9	10.2	0.0	0.0	100.0	100.0	100.0	100.0
100	25	5.4	0.4	0.0	0.0	12.5	3.3	0.0	0.0	10.5	2.6	0.0	0.0	22.3	12.3	0.0	0.0
100	50	5.2	2.3	0.0	0.0	75.5	72.4	0.0	0.0	10.6	6.0	0.0	0.0	85.8	87.1	0.0	0.1
100	100	4.9	3.9	0.0	0.0	100.0	100.0	57.0	57.5	10.6	8.9	0.0	0.0	100.0	100.0	73.7	73.0
100	250	5.2	4.5	0.0	0.0	100.0	100.0	100.0	100.0	10.5	10.3	0.0	0.0	100.0	100.0	100.0	100.0
250	25	5.0	0.0	0.0	0.0	18.6	0.8	0.0	0.0	10.1	0.4	0.0	0.0	29.2	6.9	0.0	0.0
250	50	5.0	1.0	0.0	0.0	97.8	96.1	0.0	0.0	10.7	4.4	0.0	0.0	99.2	98.9	0.0	0.0
250	100	5.0	3.4	0.0	0.0	100.0	100.0	84.4	87.1	10.1	7.9	0.0	0.0	100.0	100.0	93.4	94.0
250	250	5.0	4.8	0.0	0.0	100.0	100.0	100.0	100.0	10.0	9.3	0.0	0.0	100.0	100.0	100.0	100.0
Late positive variance shift (POS)																	
50	25	4.6	1.6	17.7	11.0	57.7	35.6	77.8	50.8	9.9	7.0	30.7	21.3	73.6	64.3	92.3	68.4
50	50	3.3	3.2	26.0	13.1	99.0	98.8	99.8	95.4	8.2	8.2	39.6	23.4	99.9	99.9	100.0	98.7
50	100	3.1	3.3	21.0	12.2	100.0	100.0	100.0	100.0	7.6	8.4	35.0	21.9	100.0	100.0	100.0	100.0
50	250	2.6	2.7	18.4	12.3	100.0	100.0	100.0	100.0	7.3	8.0	32.1	22.4	100.0	100.0	100.0	100.0
100	25	4.7	0.5	29.1	15.6	85.5	55.6	92.7	72.9	10.0	3.7	44.7	27.0	92.6	83.4	99.0	86.9
100	50	4.6	2.9	34.4	17.4	100.0	100.0	100.0	99.9	9.4	7.8	50.8	29.6	100.0	100.0	100.0	100.0
100	100	3.5	2.9	30.0	16.3	100.0	100.0	100.0	100.0	8.0	7.2	46.7	28.1	100.0	100.0	100.0	100.0
100	250	3.4	3.1	28.8	15.4	100.0	100.0	100.0	100.0	8.3	8.1	45.3	26.8	100.0	100.0	100.0	100.0
250	25	4.4	0.1	57.3	29.4	99.7	85.0	99.7	94.6	10.4	1.1	72.3	44.1	99.9	98.3	100.0	98.8
250	50	3.4	1.9	56.0	27.7	100.0	100.0	100.0	100.0	9.0	6.3	73.7	42.0	100.0	100.0	100.0	100.0
250	100	4.0	3.7	53.7	26.8	100.0	100.0	100.0	100.0	9.0	8.7	71.8	41.7	100.0	100.0	100.0	100.0
250	250	3.1	3.3	57.3	29.1	100.0	100.0	100.0	100.0	7.4	8.1	72.3	44.0	100.0	100.0	100.0	100.0

Notes: τ, $\hat{\tau}$, HS and DH refer to the PURT statistics given in (12), (16), (2) and (3) respectively. Power is not size adjusted. All results are based on 5000 replications. Data is generated according to DGP1 in (18) and all tests are computed on detrended data.

covariance matrices induces considerable power loss under a small time dimension $T = 25$. However, this power loss vanishes with increasing T. Furthermore, the new test $\hat{\tau}$ is generally as powerful as t_{HS} and more powerful than t_{DH}. Hence, it is worthwhile noting that our adjustment for obtaining robustness to time-varying volatility does not come at a cost of reduced power. In view of the fact that the reported empirical powers are not size adjusted, the power estimates for $\hat{\tau}$ are rather remarkable.

When early negative variance breaks are introduced, t_{HS} and t_{DH} display zero rejection frequencies under the null hypothesis. On the contrary, $\hat{\tau}$ holds remarkable size control, except for small T ($T = 25$) where it is substantially undersized. These size distortions, however, improve markedly as the time dimension increases to $T = 50$. The new test also has significant power under early variance breaks although it is less than the power under homoskedasticity. In comparison with $\hat{\tau}$, the White-type tests t_{HS} and t_{DH} have substantially weaker power, with both tests showing almost zero probability of rejecting the alternative hypothesis until the time dimension increases to $T = 100$.

Size distortions of t_{HS} and t_{DH} are also observed when a late positive volatility shift is considered, but this time with huge oversizings. On the contrary, $\hat{\tau}$ displays a fairly good size precision. Consistent with results in Herwartz et al. (2016) for non-trending data, power seems to be unaffected by late positive variance breaks but reduced by early negative volatility shifts. In general, simulation results documented in Table 1 demonstrate not only the risk of using t_{HS} and t_{DH} for trending time series, but also the satisfactory finite sample performance of $\hat{\tau}$ for trending heteroskedastic data.

5.2.2 DGPs with cross-sectionally correlated panels

The left-hand side block of Table 2 documents simulation results for $\hat{\tau}$ applied on data generated according to DGP1 for weakly correlated panels. Results available upon request show that size distortions of t_{HS} and t_{DH} observed for cross-sectionally independent panels (Table 1) carry over to panels with weak forms of cross-sectional correlation. Hence, we focus on the implications of cross-sectional correlation for the new test $\hat{\tau}$. Confirming the asymptotic considerations, a relatively larger cross-sectional dimension N is required for the empirical size of $\hat{\tau}$ to come closer to the nominal significance levels. Moreover, the statistic $\hat{\tau}$ is less powerful under the SAR(1) model than under independent panels—a result consistent with those documented in Herwartz et al. (2016) for non-trending series.

Table 2: Empirical rejection frequencies of $\hat{\tau}$, diverse scenarios

		DGP1, SAR(1) model				DGP2, Independence				DGP2, SAR(1) model			
		5%		10%		5%		10%		5%		10%	
N	T	size	power	size	power	size	power	size	power	size	power	size	power
Constant variance (HOM)													
50	25	4.3	16.2	11.2	28.9	0.0	0.0	0.0	0.1	0.0	0.1	0.5	0.7
50	50	3.8	60.4	10.3	78.0	0.1	29.7	0.8	50.3	0.2	8.1	1.2	20.6
50	100	3.6	100.0	9.3	100.0	1.3	100.0	3.7	100.0	1.2	95.4	4.1	98.7
50	250	3.2	100.0	8.6	100.0	2.9	100.0	6.7	100.0	1.9	100.0	5.5	100.0
100	25	4.8	24.7	10.6	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2
100	50	4.4	89.7	10.5	95.6	0.1	58.7	0.3	77.6	0.1	17.7	0.9	36.8
100	100	4.2	100.0	10.0	100.0	1.0	100.0	3.0	100.0	1.0	100.0	4.0	100.0
100	250	3.8	100.0	9.2	100.0	2.5	100.0	6.0	100.0	2.2	100.0	6.4	100.0
250	25	3.7	42.6	8.9	61.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
250	50	4.5	99.9	9.9	100.0	0.0	94.0	0.1	98.3	0.1	48.1	0.6	68.4
250	100	4.1	100.0	9.1	100.0	0.4	100.0	1.3	100.0	0.8	100.0	2.9	100.0
250	250	4.5	100.0	10.2	100.0	2.6	100.0	5.9	100.0	2.5	100.0	6.4	100.0
Early negative variance shift (NEG)													
50	25	2.9	5.1	7.7	13.7	0.0	0.0	0.0	0.0	0.1	0.0	0.5	0.3
50	50	3.3	20.8	8.9	38.1	0.1	0.4	0.4	2.0	0.3	0.9	1.4	3.0
50	100	3.7	79.2	9.6	90.4	0.7	87.7	3.0	95.9	0.8	30.3	3.1	55.3
50	250	3.5	100	8.5	100.0	3.0	100.0	8.0	100.0	1.7	100.0	6.3	100.0
100	25	2.0	4.8	5.7	13.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
100	50	3.3	35.3	8.5	53.4	0.0	0.2	0.1	0.7	0.2	0.5	0.6	2.1
100	100	4.6	97.7	10.1	99.3	0.6	99.6	2.0	100.0	0.9	66.3	2.9	84.8
100	250	4.4	100.0	9.5	100.0	3.7	100.0	8.5	100.0	2.3	100.0	6.0	100.0
250	25	0.6	3.6	3.1	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
250	50	3.0	61.9	6.9	78.1	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.9
250	100	4.2	100.0	9.2	100.0	0.2	100.0	1.2	100.0	0.5	97.9	2.0	99.5
250	250	4.6	100.0	9.8	100.0	3.9	100.0	9.5	100.0	2.5	100.0	7.6	100.0
Late positive variance shift (POS)													
50	25	1.8	14.8	7.9	34.7	0.0	0.3	0.4	4.4	0.2	0.7	1.5	4.5
50	50	1.8	56.7	7.1	80.1	0.8	63.1	4.7	83.6	0.6	16.4	3.8	37.9
50	100	1.4	99.0	6.0	99.8	2.2	100.0	6.6	100.0	0.9	85.8	4.8	96.8
50	250	1.1	100.0	5.6	100.0	2.8	100.0	7.1	100.0	0.9	100.0	4.9	100.0
100	25	2.0	24.8	7.2	50.0	0.0	0.2	0.2	3.2	0.1	0.5	0.8	4.4
100	50	2.6	89.8	7.6	97.0	1.2	91.9	4.4	98.4	1.0	39.9	4.1	66.0
100	100	2.6	100.0	7.7	100.0	3.0	100.0	8.2	100.0	1.6	99.7	5.8	100.0
100	250	2.2	100.0	7.0	100.0	3.1	100.0	8.3	100.0	2.0	100.0	6.0	100.0
250	25	0.8	48.1	4.3	77.4	0.0	0.0	0.0	1.1	0.0	0.2	0.1	2.9
250	50	3.1	99.9	8.0	100.0	0.4	100.0	3.7	100.0	0.8	85.0	4.2	95.7
250	100	3.0	100.0	8.2	100.0	3.1	100.0	9.0	100.0	2.1	100.0	6.3	100.0
250	250	2.8	100.0	7.7	100.0	4.1	100.0	10.1	100.0	2.5	100.0	7.8	100.0

Notes: Data is generated according to DGP1 in (18) for results in the left-hand side block, while DGP2 in (19) is used to generate data for results documented in the middle and right-hand side blocks of the table. Testing is performed on detrended data. For DGP2, detrending is preceded by prewhitening. Power is not size adjusted and all results are based on 5000 replications.

5.2.3 DGPs with serially correlated innovations

To evaluate how the proposed test $\hat{\tau}$ performs for data with serially correlated disturbances, we generate data according to DGP2 in (19) and subject it to prewhitening before detrending. The corresponding simulation results are documented in the middle- and right-hand side blocks of Table 2. The results show that serial correlation and the ensuing prewhitening procedure entail marked size distortions for small time dimensions. This result could be explained by noting that estimation errors arising from the prewhitening procedure introduce finite sample correlations between the lagged level and first differenced series, thereby inducing a non-zero mean to the numerator of the test statistic in (16). However, size distortions vanish as T grows, and empirical power grows in T and N.

5.2.4 Summary of simulation results

The simulation results reported in Table 1 show that existing heteroskedasticity-robust PURTs exhibit huge size distortions (either undersizing or oversizing) when applied to detrended data with time-varying volatility. The proposed test, however, performs remarkably well in this scenario. Results documented in Table 2 show that the new test has fairly good finite sample properties even when the data are not only trending and heteroskedastic, but also cross-sectionally and serially correlated. Therefore, the new test should be helpful in (often complex) empirical applications. However, results not reported here for space considerations show that $\hat{\tau}$ does not remain pivotal under strong forms of cross-sectional dependence such as factor structures (Pesaran, 2007). An effective way of panel unit root testing under strong forms of cross-sectional correlation is to remove the common factor from the data (see for example Bai and Ng, 2004 and Moon and Perron, 2004). While the test in Westerlund (2014) uses this approach, it is, however, not pivotal in the presence of linear trends.

6 Is energy use per capita trend or difference stationary?

6.1 Background

Whether energy use per capita is trend or difference stationary has been intensively investigated in the past two decades. The growing interest in testing the stationarity of per capita energy consumption is attributed to three main reasons (e.g., Hsu et al., 2008; Narayan and Smyth, 2007). First, knowing the direction of causality between per capita energy use and economic growth has gained significant policy relevance as it has direct implications on governments' involvement in global efforts to reduce greenhouse gas emissions. On the one hand, if causality runs from energy consumption to growth, reductions in energy use will have adverse effects on economic growth and, hence, generates reluctance on the part of policy makers to commit to substantial energy use reductions. On the other hand, if causality runs from growth to energy use, and not vice versa, reductions in energy consumption will not be harmful for economic growth. The order of integration of energy use per capita has implications on testing and interpreting the (causal) relationship between energy use and GDP per capita. For instance, Granger causality tests employing level vector autoregressions could be misleading if the series are nonstationary and not cointegrated. Conversely, Granger causality testing by means of variables in levels will be appropriate if the series are either stationary or cointegrated. Consequently, unit root testing is routinely performed before testing for cointegration between energy use and GDP per capita.

Second, stationarity of energy use per capita has implications for the effectiveness of energy policies such as import tariffs on fuels and vehicles or carbon taxes on transportation fuels. In particular, if energy consumption is a stationary process, it will return to its trend after a policy shock. This implies that energy saving policies will have transitory effects only. On the other hand, if energy consumption contains a unit root, such policies will have a permanent impact. Furthermore, nonstationarity implies that (permanent) shocks to energy use are more likely to affect other sectors of the economy as well as macroeconomic aggregates (Narayan and Smyth, 2007).

Third, the order of integration of energy consumption has implications for forecasting energy demand. For instance, if energy consumption is trend stationary, its past behaviour offers valuable information to forecast future energy demand. However, if energy consumption is a unit root process, it does not follow a predictable path and, hence, forecasting energy demand will be more difficult than in the stationary case.

Efforts to test for a unit root in energy use per capita have initially relied on univariate tests.[4] Most of these studies, including Glasure and Lee (1998), Beenstock et al. (1999) and McAvinchey and Yannopoulos (2003) report that the null hypothesis of an I(1) energy consumption series can not be rejected at conventional levels of significance. As an exception to this general conclusion, Altinay and Karagol (2004) document evidence in favor of characterizing energy use in Turkey during 1950–2000 as a trend stationary process. However, given the low power of univariate tests in finite samples, it is not clear if the failure to reject the null of a unit root is an evidence of a truly I(1) series. To circumvent this problem, a few studies have recently applied PURTs to examine the stationarity of energy use per capita. Results have been generally mixed, however. For instance, Joyeux and Ripple (2007) employ the PURTs suggested in Levin et al. (2002) and Im et al. (2003) and find that energy consumption measures are I(1). Narayan and Smyth (2007), on the other hand, report that the unit root null hypothesis can be rejected at the 10% level of significance for 56 of the 182 countries they considered. However, they find strong evidence of a (trend) stationary energy consumption by employing the PURT of Im et al. (2003). Nevertheless, these results should be seen with caution as the studies employ standard PURTs, which are not pivotal if the series exhibit volatility shifts.

6.2 Panel unit root test results

In this section, we study the order of integration of energy use per capita using the heteroskedasticity-robust test suggested in this paper, $\hat{\tau}$, vis-a-vis heteroskedasticity-robust tests of Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2012a). We analyse annual data of energy use per capita (kilogram of oil equivalent per capita) obtained from World Development Indicators.[5] In this data set, energy use refers to *"use of primary energy before transformation to other end-use fuels, which is equal to*

[4] See Hsu et al. (2008) for a review of the empirical literature on unit root testing of energy use per capita.

[5] `www.data.worldbank.org`. Accessed on September 23, 2016.

indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport." The study covers 23 OECD economies that are selected according to data availability, from 1960 to 2014.[6] As transforming the series into natural logarithms before undertaking unit root testing is a standard practice in the literature, we test for unit roots both on original series as well as their logarithmic values.

To get an impression if variances in the energy use per capita series exhibit significant changes over time, we plot variance profiles in Figure 1. Variance profiles $\hat{\vartheta}_i(w)$ are computed as

$$\hat{\vartheta}_i(w) = \frac{\sum_{t=1}^{\lfloor sT \rfloor} \hat{\eta}_{it}^2 + (wT - \lfloor wT \rfloor)\hat{\eta}_{i\lfloor wT \rfloor+1}^2}{\sum_{t=1}^{T} \hat{\eta}_{it}^2}, 0 \leq w \leq 1, \tag{20}$$

where the $\hat{\eta}_{it}$'s are obtained as residuals from AR(1) regressions of the series. Plotting $\hat{\vartheta}_i(w)$ against w, it is straightforward to see that a homoskedastic series would fall on the 45° line and deviations from the diagonal indicate time varying variances. Figure 1 reveals that time-varying variances characterize energy per capita series in most cross section members.

Panel unit root test results are reported in Table 3. Results for all the tests overwhelmingly show that energy use per capita has a unit root. This evidence is consistent with the findings of most of the empirical studies on the area, except, e.g., Narayan and Smyth (2007). However, it is well-known that unit root test results often depend on the specific time period chosen for study. To address this caveat, we perform panel unit root testing on rolling windows of 40 years. Corresponding results depicted in Figure 2 show that while energy use per capita is difference stationary for most of the period, it could be considered trend stationary—at least at the 10 percent significance level—for the sample periods starting between 1965 and 1968. It is worthwhile noting that $\hat{\tau}$ has the lowest $p-$value of the three tests in almost all the considered periods and could suggest an inferential outcome which is distinct from that of the other two tests. In particular, for the period spanning 1966-2005 and based on the 5 percent significance level, $\hat{\tau}$ implies that log energy per capita series can be considered trend stationary while the other two tests suggest to treat the series as difference stationary. Moreover, our

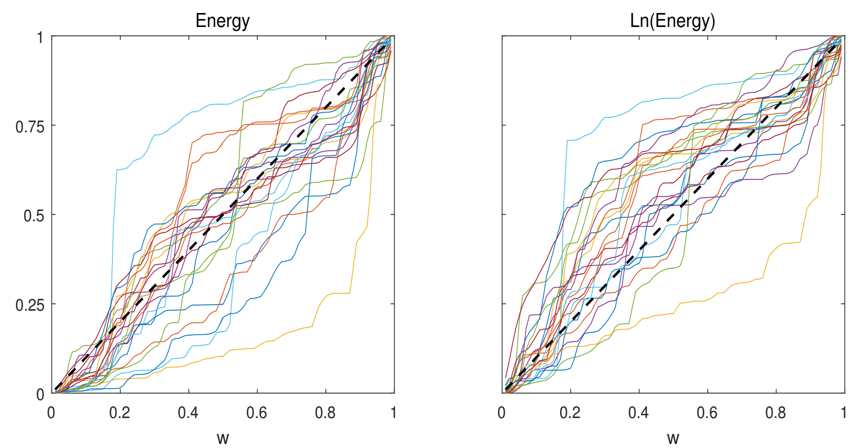
[6]The economies are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States.

Table 3: Is energy use per capita trend or difference stationary?

Period	Energy use p.c.						Ln (energy use p.c.)					
	y			Δy			y			Δy		
	$\hat{\tau}$	HS	DH	$\hat{\tau}$	HS	DH	$\hat{\tau}$	HS	DH	$\hat{\tau}$	HS	DH
Full period												
1960-2014	0.71	0.55	1.36	**-2.79**	**-3.18**	**-2.56**	1.56	1.24	1.37	**-2.86**	**-2.88**	**-2.64**
50 years window												
1960-2009	0.46	0.43	1.24	**-2.65**	**-2.82**	**-2.17**	0.86	0.91	1.39	**-2.64**	**-2.40**	**-2.13**
1961-2010	-0.18	-0.01	0.93	**-2.82**	**-2.81**	**-2.21**	0.77	0.84	1.34	**-2.81**	**-2.59**	**-2.02**
1962-2011	-0.18	0.00	0.40	**-2.71**	**-2.81**	**-2.32**	0.73	0.83	1.04	**-2.69**	**-2.56**	**-2.01**
1963-2012	-0.29	-0.10	0.18	**-2.69**	**-2.85**	**-2.57**	0.28	0.55	0.64	**-2.76**	**-2.66**	**-2.17**
1964-2013	-0.57	-0.36	-0.31	**-2.67**	**-3.25**	**-2.61**	0.04	0.39	-0.33	**-2.84**	**-2.91**	**-2.74**
1965-2014	-0.27	-0.14	-0.66	**-2.70**	**-3.11**	**-2.18**	0.19	0.48	-0.85	**-2.79**	**-2.83**	**-2.51**

Notes: Reported numbers are estimates of the panel unit root tests $\hat{\tau}$, t_{HS} and t_{DH}. Testing is performed on data that is first prewhitened and then recursively detrended. The lag order used for prewhitening is selected based on the AIC criterion, with the maximum lag lenth set to two. 'Ln' denotes the natural logarithmic transformation. Bold entries represent cases in which the panel unit root null hypothesis is rejected with 5% significance.

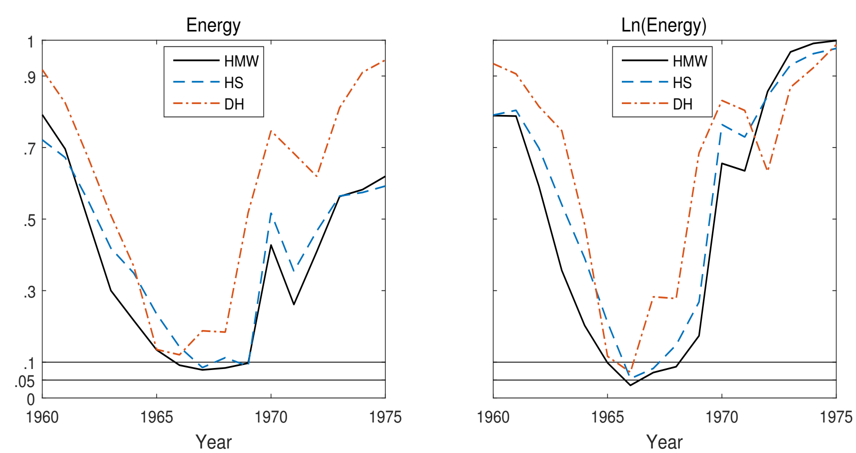
Figure 1: Estimated variance profiles



Notes: 'Ln' denotes the natural logarithmic transformation.

results also highlight the risk of deciding on stationarity of series using one specific time window.

Figure 2: Panel unit root testing over 40-years windows



Notes: The figures depict $p-$values from the panel unit root tests $\hat{\tau}$ (HMW), t_{HS} and t_{DH}. 'Year' represents the year at which the 40-years sample period begins. For further notes, see Table 3.

7 Conclusions

In this paper, we suggested a new panel unit root test (PURT) that works well when the series are trending and exhibit time-varying volatility. The test makes use of the recursive detrending scheme suggested in Demetrescu and Hanck (2014), and the construction of the test statistic fully accounts for non-zero expectation of the pooled panel regression estimator and the variance of its centered counterpart. Accordingly, the resulting test statistic has a Gaussian limiting distribution. Monte Carlo simulation results show that the test has satisfactory finite sample properties. In particular, the test tends to be conservative, while it shows remarkable power. Hence, this test should be useful in panel unit root testing of several trending macroeconomic and financial time series such as GDP per capita, industrial production, money supply and commodity prices.

The empirical illustration examined the order of integration of energy use per capita. Results using data from 23 OECD economies for the period 1960-2014 show that energy use per capita is often difference stationary. Yet, there are also a few sub-periods for which the series could be considered as trend stationary.

A particular limitation of the suggested test is that it does not perform well under a strong form of cross-sectional dependence. An effective way of panel unit root testing under strong forms of cross-sectional correlation is to remove the common factor from the data (Bai and Ng, 2004; Moon and Perron, 2004). Consequently, it appears worthwhile to see in a future research if such an approach would yield a panel unit root test that works for strongly correlated panels with trending and heteroskedastic time series.

Acknowledgements

We thank Jörg Breitung and Matei Demetrescu for helpful comments and suggestions.

Appendix

In order to prove Proposition 1 we proceed in three steps. First, stating Lemmas 1 and 2 below we are explicit on the order properties of the variance s^2_{NT} in (13) and define a mixing array which is essential to prove the asymptotic result for our test statistic (Part A.1). Second, before we derive asymptotic normality for $\hat{\tau}$ defined in (16), we establish a corresponding result for τ assuming that time specific expectations and variances are known (A.2). Third, we discuss the stochastic properties of the estimated moments $\hat{\nu}_t$ and $\hat{s}^2_{NT}$ and build upon the result for τ to finally derive the Gaussian limit distribution for $\hat{\tau}$ and thus, to prove Proposition 1 (A.3). The following derivations proceed under the null hypothesis and assumptions $\mathcal{A}$. Furthermore, we assume $N/T^2 \to 0$ throughout.

A.1 - Variance order and mixing array

Recalling from Section 4, the detrending scheme in (7) and (8) obtains coefficients $a_{i,t-1}$, finite for all $i < t$ and $t \leq T$, i.e.,

$$a_{i,t-1} = 1 + \frac{2}{t-1}(t-i) - 3\left(1 - \frac{(i-1)i}{(t-1)t}\right).$$

Let $\tilde{a}_{i,t-1} = \left(1 - \frac{1}{T}\right) a_{i,t-1}$ and $\bar{a}_{i,t-1} = \frac{1}{T} a_{i,t-1}$. The mean of $\tilde{\boldsymbol{y}}'_{t-1}\Delta\boldsymbol{y}^*_t$ is

$$\begin{aligned}
\nu_t = E\left[\tilde{\boldsymbol{y}}'_{t-1}\Delta\boldsymbol{y}^*_t\right] &= E\left[\sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}'_i\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}'_i\boldsymbol{e}_k\right)\right] \\
&= -\sum_{i=1}^{t-1}\bar{a}_{i,t-1}E\left[\boldsymbol{e}'_i\boldsymbol{e}_i\right] \\
&= -\sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i), \qquad (21)
\end{aligned}$$

since $E[\boldsymbol{e}'_i\boldsymbol{e}_k] = 0$ for all $i \neq k$. For the variance, we have

$$s^2_{NT} = \frac{1}{NT}\left(E\left[\sum_{t=2}^{T}\tilde{\boldsymbol{y}}'_{t-1}\Delta\boldsymbol{y}^*_t\right]^2 - \left(\sum_{t=2}^{T}\nu_t\right)^2\right) = \zeta_1 - \zeta_2 + \zeta_3 + \zeta_4 + \zeta_5 - \frac{1}{NT}\left(\sum_{t=2}^{T}\nu_t\right)^2, \qquad (22)$$

where the sums ζ_i, $i = 1, \ldots, 5$, are defined as in (14). Since

$$\frac{1}{NT}\left(\sum_{t=2}^{T}\nu_t\right)^2 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\mathrm{tr}(\Omega_i)^2 + \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T-1}\sum_{s=t+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\mathrm{tr}(\Omega_i)^2$$
$$+ \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}\mathrm{tr}(\Omega_i)\mathrm{tr}(\Omega_j)$$

we can rewrite s_{NT}^2 in (22) as

$$s_{NT}^2 = \tilde{\zeta}_1 - \zeta_2 + \zeta_3 + \tilde{\zeta}_4 + \tilde{\zeta}_5, \qquad (23)$$

where ζ_2 and ζ_3 are defined in (14) and

$$\tilde{\zeta}_1 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}\mathrm{tr}(\Omega_i\Omega_j)$$

$$\tilde{\zeta}_4 = \frac{1}{NT}\sum_{i=1}^{T}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \mathrm{tr}(\Omega_i)^2 + \sum_{j=1\neq i,t}^{T}\mathrm{tr}(\Omega_i\Omega_j)\right)$$
$$= \frac{1}{NT}\sum_{i=1}^{T}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \mathrm{tr}(\Omega_i)^2\right) + \frac{1}{NT}\sum_{i=1}^{T}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\sum_{j=1\neq i,t}^{T}\mathrm{tr}(\Omega_i\Omega_j)$$
$$= \tilde{\zeta}_{41} + \tilde{\zeta}_{42}$$

$$\tilde{\zeta}_5 = \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \mathrm{tr}(\Omega_i)^2 + \sum_{j=1\neq i,t,s}^{T}\mathrm{tr}(\Omega_i\Omega_j)\right)$$
$$= \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \mathrm{tr}(\Omega_i)^2\right)$$
$$+ \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\sum_{j=1\neq i,t,s}^{T}\mathrm{tr}(\Omega_i\Omega_j)$$
$$= \tilde{\zeta}_{51} + \tilde{\zeta}_{52}.$$

The following lemma characterizes the variance in more detail.

Lemma 1. *Under assumptions $\mathcal{A}$ the variance s_{NT}^2 is of order $\mathcal{O}(T)$. Moreover, $s_{NT}^2/T > 0$ for all $N, T \geq 1$.*

Proof. First, we determine the order of s_{NT}^2 as provided in (23). Then, for instance,

$$\frac{\tilde{\zeta}_1}{T} = \frac{2}{NT^4}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}a_{i,s-1}a_{j,t-1}\mathrm{tr}(\Omega_i\Omega_j)$$

is bounded in T and in N because $\mathrm{tr}(\Omega_i\Omega_j) = \mathcal{O}(N)$, i.e. $\tilde{\zeta}_1 = \mathcal{O}(T)$. Analogously, it follows $\zeta_2 = \mathcal{O}(T)$, $\zeta_3 = \mathcal{O}(T)$, $\tilde{\zeta}_{42} = \mathcal{O}(1)$ and $\tilde{\zeta}_{52} = \mathcal{O}(T)$. Furthermore, since

22

$E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \text{tr}(\Omega_i)^2 = \mathcal{O}(N)$, one has $\tilde{\zeta}_{41} = \mathcal{O}(T^{-1})$ and $\tilde{\zeta}_{51} = \mathcal{O}(1)$. Altogether, $s_{NT}^2 = \mathcal{O}(T)$.

Secondly, the variance is greater or equal to zero, $s_{NT}^2 \geq 0$, by definition (cf. equation (22)). To see that s_{NT}^2/T strictly exceeds zero for all $N, T \geq 1$, the variance of the numerator

$$\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} \left(a_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \frac{1}{T} a_{i,t-1} \sum_{k=2}^{T} \boldsymbol{e}_i'\boldsymbol{e}_k \right) = \underbrace{\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} a_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t}_{=:\mathcal{X}_1} - \underbrace{\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} \frac{1}{T} a_{i,t-1} \sum_{k=2}^{T} \boldsymbol{e}_i'\boldsymbol{e}_k}_{=:\mathcal{X}_2}$$

can be rewritten as $s_{NT}^2 = \text{Var}[\mathcal{X}_1 - \mathcal{X}_2] = \text{Var}[\mathcal{X}_1] + \text{Var}[\mathcal{X}_2] - 2 \cdot \text{Cov}[\mathcal{X}_1, \mathcal{X}_2]$. The components of $\text{Var}[\mathcal{X}_1]$ consist of terms $\text{tr}(\Omega_i\Omega_t)/N$ which are strictly positive, $\text{tr}(\Omega_i\Omega_j)/N \geq \lambda_i^{(1)}(\sum_{l=1}^{N} \lambda_j^{(l)})/N > 0$ for all $i, j = 1, \ldots, T$, $N \geq 1$, and eigenvalues $\lambda_i^{(1)}, \lambda_j^{(l)} > 0$ (from assumption $\mathcal{A}(ii)$). Hence, it can be shown that $\text{Var}[\mathcal{X}_1]/T > 0$ for all $N, T \geq 1$. Furthermore, the variance terms $\text{Var}[\mathcal{X}_1] + \text{Var}[\mathcal{X}_2]$ can be shown to dominate the covariance term $2 \cdot \text{Cov}[\mathcal{X}_1, \mathcal{X}_2]$ so that $s_{NT}^2/T = \text{Var}[\mathcal{X}_1 - \mathcal{X}_2]/T > 0$ for all $N, T \geq 1$. □

To show the asymptotic normality of $\hat{\tau}$ in (16), we employ a central limit theorem for near-epoch dependent sequences. For this, we define a mixing array[7]

$$\boldsymbol{V}_{T,t} = \left(\boldsymbol{e}_t, \sum_{k=t+1}^{T} \boldsymbol{e}_k \right). \tag{24}$$

The generated sigma algebra corresponds to

$$\mathcal{F}_{T,t-m}^{t+m} = \sigma(\boldsymbol{V}_{T,s}, t-m \leq s \leq t+m) = \sigma(\boldsymbol{e}_{t-m}, \ldots, \boldsymbol{e}_{t+m}, \mathcal{E}_{T,t-m}, \ldots, \mathcal{E}_{T,t+m}),$$

where $\mathcal{E}_{T,t+m} := \sum_{k=t+m+1}^{T} \boldsymbol{e}_k$. In particular, $\mathcal{F}_{T,-\infty}^{t} = \sigma(\ldots, \boldsymbol{e}_t, \ldots, \mathcal{E}_{T,t})$. This definition of the sigma algebra is similar to the one used in Lemma 3 of Demetrescu and Hanck (2014), but contains the vector $(\boldsymbol{e}_t, \mathcal{E}_{T,t})$ instead of the sum of the two entries. Using the notation of Davidson (1994) we state the following result:

Lemma 2. *$\boldsymbol{V}_{T,t}$ in (24) is α-mixing of size $-\beta$ for $0 \leq \beta < \infty$.*

[7] For simplicity the subscript N is omitted here, since the process is near-epoch dependent with respect to the time dimension.

Proof. To show the mixing property of $\boldsymbol{V}_{T,t}$ consider the sequence

$$\alpha_m = \sup_t \sup_{A \in \mathcal{F}^{\infty}_{T,t+m}, B \in \mathcal{F}^{t}_{T,-\infty}} \big| P(A \cap B) - P(A)P(B) \big|$$

for all $T \geq 1$ and events A and B. The second supremum is taken with respect to the sigma algebras $\mathcal{F}^t_{T,-\infty} = \sigma(\ldots, \boldsymbol{e}_t, \ldots, \mathcal{E}_{T,t})$ and $\mathcal{F}^{\infty}_{T,t+m} = \sigma(\boldsymbol{e}_{t+m}, \ldots, \mathcal{E}_{T,t+m}, \ldots)$. Noticing that the $\boldsymbol{e}_i$'s are uncorrelated, dependence between A and B (i.e., $|P(A \cap B) - P(A)P(B)\big| > 0$) can only occur by involving terms of $\mathcal{E}_{T,t}$. More precisely, the sums $\mathcal{E}_{T,t} = \sum_{k=t}^{T} \boldsymbol{e}_k$ and $\mathcal{E}_{T,t+m} = \sum_{k=t+m+1}^{T} \boldsymbol{e}_k$ both include error terms $\{\boldsymbol{e}_{t+m+1}, \ldots, \boldsymbol{e}_T\}$ such that $\alpha_m \neq 0$ for events

$$A, B \in \mathcal{F}^t_{T,-\infty} \cap \mathcal{F}^{\infty}_{T,t+m} = \sigma\left(\mathcal{E}_{T,t+m}, \mathcal{E}_{T,t+m+1}, \ldots\right) \subseteq \sigma\left(\boldsymbol{e}_{t+m+1}, \boldsymbol{e}_{t+m+2}, \ldots, \boldsymbol{e}_T\right). \quad (25)$$

For increasing m the number of random variables generating the sigma algebra decreases. For $m > T - t - 1$ the generated sigma algebra in (25) is the empty set. Thus, $\alpha_m = 0$ for $m > T - t - 1$ for all $T \geq 1$ and $-\infty \leq t \leq \infty$. It follows $\alpha_m = \mathcal{O}(m^{-\beta})$ for all $0 \leq \beta < \infty$. □

A.2 - Asymptotic distribution with true moments

In the following, $\hat{\nu}_t$ and $\hat{s}_{NT}$ are substituted by their theoretical counterparts so that asymptotic normality of τ defined in (12) is shown first. To prove asymptotic normality of τ we rewrite the numerator from (12) as

$$\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} (\tilde{\boldsymbol{y}}'_{t-1} \Delta \boldsymbol{y}^*_t - \nu_t) = \sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \left(\sum_{i=1}^{t-1} \left(\tilde{a}_{i,t-1} \boldsymbol{e}'_i \boldsymbol{e}_t - \bar{a}_{i,t-1} \sum_{k=2, k\neq t}^{T} \boldsymbol{e}'_i \boldsymbol{e}_k \right) - \nu_t \right).$$

From the variance

$$s^2_{NT} = E\left[\left(\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} \left(\tilde{a}_{i,t-1} \boldsymbol{e}'_i \boldsymbol{e}_t - \bar{a}_{i,t-1} \sum_{k=2,k\neq t}^{T} \boldsymbol{e}'_i \boldsymbol{e}_k \right) - \nu_t \right)^2 \right],$$

a standardized sequence is given by

$$\begin{aligned} X_{NT,t} :=& \frac{1}{\sqrt{NT}} (\tilde{\boldsymbol{y}}'_{t-1} \Delta \boldsymbol{y}^*_t - \nu_t) \Big/ s_{NT} \\ =& \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} \left(\tilde{a}_{i,t-1} \boldsymbol{e}'_i \boldsymbol{e}_t - \bar{a}_{i,t-1} \sum_{k=2,k\neq t}^{T} \boldsymbol{e}'_i \boldsymbol{e}_k + \bar{a}_{i,t-1} \mathrm{tr}(\Omega_i) \right) \Big/ s_{NT}. \end{aligned} \quad (26)$$

A central limit theorem (CLT) for $\tau = \sum_{t=2}^{T} X_{NT,t}$ that controls for near-epoch dependence (NED) of $X_{NT,t}$ holds if the following conditions of Corollary 24.7 in Davidson (1994) are fulfilled:

(a) $X_{NT,t}$ is $\mathcal{F}^t_{T,-\infty}$ measurable with $E\left[X_{NT,t}\right] = 0$ and $E\left[\left(\sum_{t=2}^T X_{NT,t}\right)^2\right] = 1$.

(b) There exists a constant array $\{c_{NT,t}\}$ such that $\sup_{T,t} ||X_{NT,t}/c_{NT,t}||_r < \infty$ for $r > 2$.

(c) $X_{NT,t}$ is L_2-NED of size -1 on $\boldsymbol{V}_{T,t}$ which is α-mixing of size $-r/(r-2)$.

(d) $\sup_T \{T(\max_{1\leq t\leq T} c_{NT,t})^2\} < \infty$.

Lemma 3. *Under assumptions $\mathcal{A}$ on the error terms and $N/T^2 \to 0$ the conditions (a)-(d) are fulfilled for the sequence $X_{NT,t}$ in (26) and the mixing process $\boldsymbol{V}_{T,t}$ in (24).*

From Corollary 24.7 in Davidson (1994) and Lemma 3 asymptotic normality of τ in (12) follows directly and can be stated as

Corollary 1. *Under assumptions $\mathcal{A}$ and $N/T^2 \to 0$,*

$$\tau = \sum_{t=2}^{T} X_{NT,t} \xrightarrow{d} \mathcal{N}(0,1), \quad N, T \to \infty.$$

Remark. *The CLT in T holds for all $N \geq 1$, in particular for $N \to \infty$. The joint limit $N, T \to \infty$, furthermore, provides convergence of the sums of $\boldsymbol{e}_i'\boldsymbol{e}_t$ and thus, ensures that the assumptions of the CLT are fulfilled. Note that we show asymptotic normality in the joint limit $N, T \to \infty$ instead of the sequential limit applying the convergence properties following, for instance, from Theorem 4.4 of Billingsley (1999).*

Proof of Lemma 3. **Condition (a):** As it is a function of measurable random variables, $X_{NT,t} = f(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_t, \mathcal{E}_t)$ is measurable with respect to $\mathcal{F}^t_{T,-\infty}$. The sequence $X_{NT,t}$ is centered and standardized such that $E[X_{NT,t}] = 0$ and $E\left[\left(\sum_{t=2}^T X_{NT,t}\right)^2\right] = 1$ follow directly.

Condition (b): Let the array of constants be equal to $\{c_{NT,t}\} = \{1/s_{NT}\}$ and set $r = 4$.

Then,

$$\begin{aligned}
\left|\left|X_{NT,t}/c_{NT,t}\right|\right|_4 &= E\left[\left|X_{NT,t}\right|^4 \Big/ c_{NT,t}^4\right]^{\frac{1}{4}} = E\left[\left|X_{NT,t}\right|^4 \cdot s_{NT}^4\right]^{\frac{1}{4}} \\
&= \left[E\Big|\frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\Big(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k + \bar{a}_{i,t-1}\mathrm{tr}(\Omega_i)\Big)\Big|^4\right]^{\frac{1}{4}} \\
&= \Big(E\Big|\frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \sum_{i=1}^{t-1}\bar{a}_{i,t-1}\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k + \sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i)\Big)\Big|^4\Big)^{\frac{1}{4}} \\
&\leq \Big(\frac{1}{N^2T^2}E\Big|\sum_{i=1}^{t-1}\tilde{a}_{i,t-1}\sum_{l=1}^{N}e_{li}e_{lt}\Big|^4\Big)^{\frac{1}{4}} + \Big(\frac{1}{N^2T^6}E\Big|\sum_{i=1}^{t-1}a_{i,t-1}\sum_{\substack{k=2\\k\neq t}}^{T}\sum_{l=1}^{N}e_{li}e_{lk}\Big|^4\Big)^{\frac{1}{4}} \\
&\quad + \Big(\frac{1}{N^2T^6}E\Big|\sum_{i=1}^{t-1}a_{i,t-1}\mathrm{tr}(\Omega_i)\Big)\Big|^4\Big)^{\frac{1}{4}} \\
&< \infty \qquad \text{for all } N, T. \qquad (27)
\end{aligned}$$

The inequality holds by virtue of the Minkowski inequality. The first part in (27) is finite with similar reasoning as in equation (19) of Herwartz et al. (2016), i.e. nonzero expectations arise only from terms involving $(e_{li}e_{lt})^4$ or $e_{li}^2e_{lj}^2e_{mt}^4$, $i \neq j$. The second and the third term contain the product of error terms from the same time period ($e_{li}e_{lk}$ with $i = k$). Thus, for finiteness we need to assume finiteness up to order eight, $E|e_{lt}|^8 < \infty$, which was formulated in assumption $\mathcal{A}(iii)$. Furthermore, noticing that N and T can be related by means of $N/T^2 \to 0$, the denominator controls for increasing N and T adequately.

Condition (c): To verify this condition, $X_{NT,t}$ is shown to be near-epoch dependent on $\boldsymbol{V}_{T,t}$ meaning that

$$\Big(E\left[X_{NT,t} - E[X_{NT,t}|\, \mathcal{F}_{T,t-m}^{t+m}]\right]^2\Big)^{1/2} \leq c_{NT,t}\rho_m, \qquad (28)$$

where ρ_m is a sequence of order $\mathcal{O}(m^{-1})$ and $c_{NT,t}$ is the positive constant defined in condition (b).

The expectation of $X_{NT,t}$ conditioned on the m neighboring sigma algebras is

$$\begin{aligned}
E\left[X_{NT,t}|\, \mathcal{F}_{T,t-m}^{t+m}\right] &= E\left[\frac{1}{\sqrt{NT}}\Big(\sum_{i=1}^{t-1}\big(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\big) - \nu_t\Big)\Big/s_{NT}\Big|\, \mathcal{F}_{T,t-m}^{t+m}\right] \\
&= \frac{1}{\sqrt{NT}}\Big(\sum_{i=1}^{t-1}E\left[\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\Big|\, \mathcal{F}_{T,t-m}^{t+m}\right] - \nu_t\Big)\Big/s_{NT}, \qquad (29)
\end{aligned}$$

26

with

$$\sum_{i=1}^{t-1} E\left[\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\Big|\, \mathcal{F}_{T,t-m}^{t+m}\right]$$
$$= \sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}E\left[\boldsymbol{e}_i'\boldsymbol{e}_t\big|\, \mathcal{F}_{T,t-m}^{t+m}\right] - \bar{a}_{i,t-1}E\left[\boldsymbol{e}_i'\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_k\Big|\, \mathcal{F}_{T,t-m}^{t+m}\right]\right)$$
$$= \sum_{i=t-m}^{t-1}\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \sum_{i=1}^{t-m-1}\tilde{a}_{i,t-1}E(\boldsymbol{e}_i'\boldsymbol{e}_i) - \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=t-m,k\neq t}^{T}\boldsymbol{e}_k. \tag{30}$$

Here, parts of the conditional expectations cancel out because of measurability or zero covariance of the corresponding random variables. Inserting (30) into (29) obtains

$$E\left[X_{NT,t}|\mathcal{F}_{T,t-m}^{t+m}\right] = \frac{1}{\sqrt{NT}}\left(\sum_{i=t-m}^{t-1}\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \sum_{i=1}^{t-m-1}\tilde{a}_{i,t-1}E(\boldsymbol{e}_i'\boldsymbol{e}_i)\right.$$
$$\left. - \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{\substack{k=t-m\\k\neq t}}^{T}\boldsymbol{e}_k + \sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i)\right)\Big/ s_{NT}$$
$$= \frac{1}{\sqrt{NT}}\sum_{i=t-m}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\right.$$
$$\left. + \bar{a}_{i,t-1}\mathrm{tr}(\Omega_i) - \bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right)\Big/ s_{NT}.$$

Hence, the condition for NED sequences in (28) is fulfilled by noticing

$$\left(E\left[X_{NT,t} - E[X_{NT,t}|\mathcal{F}_{T,t-m}^{t+m}]\right]^2\right)^{\frac{1}{2}}$$
$$= \left(\frac{1}{s_{NT}\sqrt{NT}}E\left[\sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\right) - \nu_t\right.\right.$$
$$\left.\left. - \sum_{i=t-m}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k + \bar{a}_{i,t-1}\mathrm{tr}(\Omega_i) - \bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right)\right]^2\right)^{\frac{1}{2}}$$
$$= \frac{1}{s_{NT}\sqrt{NT}}\left(E\left[\sum_{i=1}^{t-m-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\left(\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k - \mathrm{tr}(\Omega_i)\right)\right) + \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right]^2\right)^{\frac{1}{2}}$$
$$= c_{NT,t}\left(\frac{1}{NT}E\left[\sum_{i=1}^{t-m-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\left(\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k - \mathrm{tr}(\Omega_i)\right)\right) + \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right]^2\right)^{\frac{1}{2}}$$
$$=: c_{NT,t}\,\rho_m.$$

In order to show that $\rho_m = \mathcal{O}(m^{-1})$, we apply Minkowski's inequality:

$$\begin{aligned}
\rho_m =& \Bigg(\frac{1}{NT} E\Bigg[\sum_{i=1}^{t-m-1} \Big(\tilde{a}_{i,t-1} \boldsymbol{e}_i' \boldsymbol{e}_t - \bar{a}_{i,t-1} \Big(\sum_{\substack{k=2 \\ k \neq t}}^{T} \boldsymbol{e}_i' \boldsymbol{e}_k - \mathrm{tr}(\Omega_i) \Big) \Big) + \sum_{i=t-m}^{t-1} \bar{a}_{i,t-1} \boldsymbol{e}_i' \sum_{k=2}^{t-m} \boldsymbol{e}_k \Bigg]^2 \Bigg)^{\frac{1}{2}} \\
\leq& \frac{1}{\sqrt{T}} \underbrace{\left(\frac{1}{N} E \left[\sum_{i=1}^{t-m-1} \Big(\tilde{a}_{i,t-1} \boldsymbol{e}_i' \boldsymbol{e}_t - \bar{a}_{i,t-1} \Big(\sum_{k=2, k \neq t}^{T} \boldsymbol{e}_i' \boldsymbol{e}_k - \mathrm{tr}(\Omega_i) \Big) \Big) \right]^2 \right)^{\frac{1}{2}}}_{=\mathcal{O}(\sqrt{t-m-1})} \\
&+ \frac{1}{\sqrt{T}} \underbrace{\left(\frac{1}{N} E \left[\sum_{i=t-m}^{t-1} \sum_{k=2}^{t-m} \bar{a}_{i,t-1} \boldsymbol{e}_i' \boldsymbol{e}_k \right]^2 \right)^{\frac{1}{2}}}_{=\mathcal{O}((m-1)(t-m)/T)} . \qquad (31)
\end{aligned}$$

The error terms have finite fourth order moments and, hence, dividing by $\sqrt{T}$ the L_2-norms are bounded for all $m, t, T \geq 1$. Furthermore, for $m \geq t-1$ the sums are zero such that $\rho_m = 0$. Consequently, for $0 \leq \beta < \infty$ we have $m^\beta \rho_m = \mathcal{O}(1)$ if $m < t-1$, because both m and ρ_m are bounded, and if $m \geq t-1$ because $\rho_m = 0$. Thus, $\rho_m = \mathcal{O}(m^{-\beta})$ for every $0 \leq \beta < \infty$ and especially, for $\beta = 1$ such that $\rho_m = \mathcal{O}(m^{-1})$.

Furthermore, from Lemma 2 it follows that $\boldsymbol{V}_{T,t}$ is mixing of size $-\beta$ for $\beta \geq 0$. In particular, for $r = 4$ the order of convergence is $-\beta = -r/(r-2) = -2$ as considered in condition (b) of Theorem 24.6 in Davidson (1994).

Condition (d): To show that this condition holds for $c_{NT,t} = 1/s_{NT}$, notice that s_{NT}^2 is of order $\mathcal{O}(T)$ following Lemma 1. Together with $s_{NT}^2/T > 0$ this directly indicates the finiteness required by condition (d):

$$\sup_T \left\{ T \left(\max_{1 \leq t \leq T} c_{NT,t} \right)^2 \right\} = \sup_T \frac{T}{s_{NT}^2} < \infty.$$

□

A.3 - Asymptotic distribution with estimated moments

Mean estimation

The representation in (21) reduces the estimation of ν_t to the estimation of terms such as $\mathrm{tr}(\Omega_i)$ so that convergence is assured by the increasing panel and time dimensions N and T. For the model residuals evaluated under the null hypothesis $\hat{\boldsymbol{e}}_t = \Delta \boldsymbol{y}_t^*$ the estimator is explicitly given as

$$\hat{\nu}_t = -\sum_{i=1}^{t-1} \bar{a}_{i,t-1} \hat{\boldsymbol{e}}_i' \hat{\boldsymbol{e}}_i.$$

The following lemma states convergence of $\frac{1}{\sqrt{NT}}(\hat{\nu}_t - \nu_t)$ so that the theoretical counterpart ν_t can be used to prove asymptotic normality of $\hat{\tau}$.

Lemma 4. *Under assumptions $\mathcal{A}$,*

$$\frac{1}{\sqrt{NT}}(\hat{\nu}_t - \nu_t) \xrightarrow{p} 0, \quad for \quad N, T \to \infty.$$

Proof. To show the convergence in probability, we rewrite

$$\frac{1}{\sqrt{NT}}(\hat{\nu}_t - \nu_t) = \frac{1}{\sqrt{NT}} \left(\sum_{i=1}^{t-1} \frac{1}{T} a_{i,t-1} \left(\hat{\boldsymbol{e}}_i' \hat{\boldsymbol{e}}_i - E[\boldsymbol{e}_i' \boldsymbol{e}_i] \right) \right).$$

From (10) we have $\hat{\boldsymbol{e}}_i = \boldsymbol{e}_i - \frac{1}{T}\sum_{t=2}^{T} \boldsymbol{e}_t$. For finite T the variance and covariance of the estimator $\hat{\boldsymbol{e}}_i$ differ from corresponding moments of $\boldsymbol{e}_i$. However, asymptotically they are equivalent. For instance, for any $i = 1, \ldots, T$,

$$\begin{aligned}
E\left[\hat{\boldsymbol{e}}_i' \hat{\boldsymbol{e}}_i\right] &= E\left[\Big(\boldsymbol{e}_i - \frac{1}{T}\sum_{t=2}^{T} \boldsymbol{e}_t\Big)'\Big(\boldsymbol{e}_i - \frac{1}{T}\sum_{t=2}^{T} \boldsymbol{e}_t\Big)\right] \\
&= E\left[\boldsymbol{e}_i' \boldsymbol{e}_i\right] - 2E\left[\boldsymbol{e}_i' \Big(\frac{1}{T}\sum_{t=2}^{T} \boldsymbol{e}_t\Big)\right] + E\left[\Big(\frac{1}{T}\sum_{t=2}^{T} \boldsymbol{e}_t\Big)'\Big(\frac{1}{T}\sum_{t=2}^{T} \boldsymbol{e}_t\Big)\right] \\
&= \Big(1 - \frac{2}{T}\Big) E\left[\boldsymbol{e}_i' \boldsymbol{e}_i\right] + \frac{1}{T^2}\sum_{t=2}^{T} E[\boldsymbol{e}_t' \boldsymbol{e}_t] \to E\left[\boldsymbol{e}_i' \boldsymbol{e}_i\right], \quad T \to \infty. \qquad (32)
\end{aligned}$$

Similarly, the higher moments converge, i.e. $E[(\hat{\boldsymbol{e}}_i{}' \hat{\boldsymbol{e}}_i)^2] \to E[(\boldsymbol{e}_i' \boldsymbol{e}_i)^2]$ for $T \to \infty$.

Applying these results and the Markov inequality we have

$$
\begin{aligned}
P\left(\left|\frac{\hat{\nu}_t-\nu_t}{\sqrt{NT}}\right|>\varepsilon\right) &< \frac{E\left[\frac{1}{NT}\left(\hat{\nu}_t-\nu_t\right)^2\right]}{\varepsilon^2}\\
&= \frac{1}{\varepsilon^2}E\left[\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\frac{1}{T}a_{i,t-1}\left(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i]\right)\right)^2\right]\\
&= \frac{1}{\varepsilon^2 NT^3}E\left[\left(\sum_{i=1}^{t-1}a_{i,t-1}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i])\right)^2\right]\\
&= \frac{1}{\varepsilon^2 NT^3}\left[\sum_{i=1}^{t-1}a_{i,t-1}^2E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i])^2\right]\right.\\
&\quad\left.+2\sum_{i=1}^{t-1}\sum_{j=i+1}^{t-1}a_{i,t-1}a_{j,t-1}E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i])\left(\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_j-E[\boldsymbol{e}_j'\boldsymbol{e}_j]\right)\right]\right]\\
&= \frac{1}{\varepsilon^2T^3}\underbrace{\sum_{i=1}^{t-1}a_{i,t-1}^2\frac{1}{N}\left(E[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2]-(\mathrm{tr}(\Omega_i))^2\right)}_{\mathcal{O}(T)}\\
&\quad+\frac{2}{\varepsilon^2NT^3}\sum_{i=1}^{t-1}\sum_{j=i+1}^{t-1}a_{i,t-1}a_{j,t-1}\underbrace{E\left[\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_j-E[\boldsymbol{e}_i'\boldsymbol{e}_i]E[\boldsymbol{e}_j'\boldsymbol{e}_j]\right]}_{=0}\\
&\longrightarrow 0 \quad \text{for} \quad \varepsilon>0, \quad N,T\to\infty.
\end{aligned}
$$

□

Variance estimation

According to the representation of s_{NT}^2 in (23) the variance estimator is

$$\hat{s}_{NT}^2=\hat{\hat{\zeta}}_1-\hat{\zeta}_2+\hat{\zeta}_3+\hat{\hat{\zeta}}_4+\hat{\hat{\zeta}}_5, \tag{33}$$

where

$$\hat{\hat{\zeta}}_1 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j)^2$$

$$\hat{\zeta}_2 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=i+1}^{T}\tilde{a}_{i,s-1}\bar{a}_{i,t-1}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_s)^2$$

$$\hat{\zeta}_3 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2$$

$$\hat{\hat{\zeta}}_4 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\left(\underbrace{(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2-(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2}_{=0}+\sum_{j=1,j\neq i,t}^{T}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j)^2\right)$$

$$\begin{aligned}\hat{\hat{\zeta}}_5 &= \hat{\hat{\zeta}}_{51}+\hat{\hat{\zeta}}_{52}\\ &= \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\underbrace{\left((\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2-(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2\right)}_{=0}\\ &\quad+\frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\sum_{j=1,j\neq i,t,s}^{T}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j)^2.\end{aligned}$$

However, unlike $\frac{1}{\sqrt{NT}}(\hat{\nu}_t-\nu_t)\xrightarrow{p}0$, the difference $\hat{s}_{NT}^2-s_{NT}^2$ does not converge in probability. To determine the order of this difference, we consider the components in (33) separately. For $N,T\to\infty$, the orders of the differences of $\hat{\hat{\zeta}}_1$, $\hat{\zeta}_2$, $\hat{\zeta}_3$ and $\hat{\hat{\zeta}}_{52}$ from their theoretical counterparts can be derived in the same form. As an example, we consider

$$\hat{\zeta}_3-\zeta_3=\sum_{i=1}^{T-1}\frac{1}{NT}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\operatorname{tr}(\Omega_i\Omega_t)\right]. \tag{34}$$

To define the order of $\hat{\zeta}_3-\zeta_3$ we use $E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2\right]\to E\left[(\boldsymbol{e}_i'\boldsymbol{e}_t)^2\right]$ and $E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_s)^2\right]\to E\left[(\boldsymbol{e}_i'\boldsymbol{e}_t)^2(\boldsymbol{e}_i'\boldsymbol{e}_s)^2\right]$ which can be derived similarly to (32). Accordingly, the difference in

(34) has mean zero but its variance does not vanish. More precisely,

$$
\begin{aligned}
&E\left[\left(\sum_{i=1}^{T-1}\frac{1}{NT}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\mathrm{tr}(\Omega_i\Omega_t)\right]\right)^2\right]\\
&=\frac{1}{N^2T^2}\left(\sum_{i=1}^{T-1}E\left[\left(\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\mathrm{tr}(\Omega_i\Omega_t)\right]\right)^2\right]\right.\\
&\quad\left.+2\sum_{i=1}^{T-2}\sum_{j=i+1}^{T-1}E\left[\left(\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\mathrm{tr}(\Omega_i\Omega_t)\right]\right)\left(\sum_{t=j+1}^{T}\tilde{a}_{j,t-1}^2\left[(\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_t)^2-\mathrm{tr}(\Omega_j\Omega_t)\right]\right)\right]\right)\\
&=\frac{1}{N^2T^2}\left(\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^4\left[E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^4\right]-(\mathrm{tr}(\Omega_i\Omega_t))^2\right]\right.\\
&\qquad+\sum_{i=1}^{T-1}2\sum_{t=i+1}^{T}\sum_{s=t+1}^{T}\tilde{a}_{i,t-1}^2\tilde{a}_{i,s-1}^2\left[E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_s)^2\right]-\mathrm{tr}(\Omega_i\Omega_t)\mathrm{tr}(\Omega_i\Omega_s)\right]\\
&\quad\left.+2\sum_{i=1}^{T-2}\sum_{j=i+1}^{T-1}E\left[\left(\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\mathrm{tr}(\Omega_i\Omega_t)\right]\right)\left(\sum_{t=j+1}^{T}\tilde{a}_{j,t-1}^2\left[(\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_t)^2-\mathrm{tr}(\Omega_j\Omega_t)\right]\right)\right]\right)\\
&=\frac{1}{N^2T^2}\left(\mathcal{O}(NT^2)+\mathcal{O}(NT^3)+\mathcal{O}(NT^4)\right)\\
&=\mathcal{O}\left(T^2/N\right). \qquad (35)
\end{aligned}
$$

Assuming weak cross-sectional dependence the order in N follows similarly to the derivation of the order $\mathrm{tr}(\Omega_i\Omega_t)=\mathcal{O}(N)$. Consequently, $\hat{\zeta}_3-\zeta_3=\mathcal{O}_p\left(\sqrt{\mathrm{Var}[\hat{\zeta}_3]}\right)=\mathcal{O}_p(T/\sqrt{N})$. Similar arguments apply for $\tilde{\zeta}_1$, ζ_2 and $\tilde{\zeta}_{52}$. Moreover, we obtain

$$\hat{\tilde{\zeta}}_{51}-\tilde{\zeta}_{51}=\frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(\mathrm{tr}(\Omega_i)^2-E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2]\right)=\mathcal{O}_p(1),$$

because we have $\left(\mathrm{tr}(\Omega_i)^2-E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2]\right)/N=\mathcal{O}(1)$ from the proof of Lemma 1. Combining these arguments, convergence of the remaining term $\hat{\tilde{\zeta}}_4-\tilde{\zeta}_4\xrightarrow{p}0$ follows directly. By implication, $(\hat{s}_{NT}^2-s_{NT}^2)=\mathcal{O}_p(T/\sqrt{N})+\mathcal{O}_p(1)$.

Proof of Proposition 1. Asymptotic normality of $\hat{\tau}$ stated in Proposition 1 follows from the asymptotic behaviour of $\hat{\nu}$ and $\hat{s}_{NT}^2$, Corollary 1 and a Taylor approximation of $\hat{\tau}$ in the true variance s_{NT}^2. Noticing that $\frac{1}{\sqrt{NT}}(\hat{\nu}_t-\nu_t)=o_p(1)$, we define the empirical version of the test statistic $\hat{\tau}$ as a function of $\hat{s}_{NT}^2$ as

$$\hat{\tau}=\tau(\hat{s}_{NT}^2)=\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}(\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*-\nu_t)\Big/\hat{s}_{NT}.$$

Applying the first-order Taylor expansion in s^2_{NT}, asymptotic normality of $\hat{\tau}$ follows as:

$$\begin{aligned}
\hat{\tau} &= \sum_{t=2}^{T} \frac{1}{\sqrt{NT}} (\tilde{\boldsymbol{y}}'_{t-1} \Delta \boldsymbol{y}^*_t - \nu_t) \Big/ \hat{s}_{NT} = \tau(s^2_{NT} + (\hat{s}^2_{NT} - s^2_{NT})) \\
&= \tau(s^2_{NT}) + (\hat{s}^2_{NT} - s^2_{NT}) \cdot \frac{\partial \tau}{\partial (s^2_{NT})} + o_p(1) \\
&= \tau(s^2_{NT}) + (\hat{s}^2_{NT} - s^2_{NT}) \left(-\frac{1}{2} \sum_{t=2}^{T} \frac{1}{\sqrt{NT}} (\tilde{\boldsymbol{y}}'_{t-1} \Delta \boldsymbol{y}^*_t - \nu_t) \right) \cdot (s^2_{NT})^{-3/2} + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0,1) + \left(\mathcal{O}_p \left(T/\sqrt{N} \right) + \mathcal{O}_p(1) \right) \mathcal{O}_p \left(\sqrt{T} \right) (\mathcal{O}(T))^{-3/2} + o_p(1), \\
&= \mathcal{N}(0,1) + o_p(1), \quad \text{for} \quad N, T \to \infty.
\end{aligned}$$

Convergence of the first term to the standard normal distribution is stated in Corollary 1, and, hence Proposition 1 follows.[8]

□

[8] To see that the remainder term is $o_p(1)$, consider, for instance, the expansion of second order $\left(\hat{s}^2_{NT} - s^2_{NT}\right)^2 \cdot \frac{\partial \tau}{\partial^2 (s^2_{NT})} = \left(\mathcal{O}_p \left(T/\sqrt{N} \right) + \mathcal{O}_p(1) \right)^2 \mathcal{O}_p(\sqrt{T})(\mathcal{O}(T))^{-5/2} = o_p(1)$ for $N, T \to \infty$.

References

Altinay, G., Karagol, E., 2004. Structural break, unit root, and the causality between energy consumption and GDP in Turkey. Energy Economics 26 (6), 985–994.

Anselin, L., 2013. Spatial econometrics: methods and models. Vol. 4. Springer Science & Business Media.

Bai, J., Ng, S., 2004. A panic attack on unit roots and cointegration. Econometrica 72 (4), 1127–1178.

Beenstock, M., Goldin, E., Nabot, D., April 1999. The demand for electricity in Israel. Energy Economics 21 (2), 168–183.

Billingsley, P., 1999. Convergence of probability measures, 2nd Edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, a Wiley-Interscience Publication.

Breitung, J., Das, S., 2005. Panel unit root tests under cross sectional dependence. Statistica Neerlandica 59 (4), 414–433.

Cavaliere, G., 2004. Unit root tests under time varying variances. Econometric Reviews 23 (4), 259–292.

Cavaliere, G., Taylor, A. M. R., 2007. Testing for unit roots in time series with non-stationary volatility. Journal of Econometrics 140 (2), 919–947.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

Demetrescu, M., Hanck, C., 2012a. A simple nonstationary-volatility robust panel unit root test. Economics Letters 117 (2), 10–13.

Demetrescu, M., Hanck, C., 2012b. Unit root testing in heteroscedastic panels using the Cauchy estimator. Journal of Business and Economic Statistics 30 (2), 256–264.

Demetrescu, M., Hanck, C., 2014. Robust inference for near-unit root processes with time-varying error variances. Econometric Reviews (forthcoming).

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

Gadea-Rivas, M. D., Gómez-Loscos, A., Pérez-Quirós, G., 2014. The two greatest. Great recession vs. great moderation. Banco de Espana Working Papers 1423, Banco de Espana.

Glasure, Y. U., Lee, A.-R., 1998. Cointegration, error-correction, and the relationship between GDP and energy: The case of South Korea and Singapore. Resource and Energy Economics 20 (1), 17–25.

Hamori, S., Tokihisa, A., 1997. Testing for a unit root in the presence of a variance shift. Economics Letters 57 (3), 245–253.

Herwartz, H., Siedenburg, F., 2008. Homogenous panel unit root tests under cross sectional dependence: Finite sample modifications and the wild bootstrap. Computational Statistics and Data Analysis 53 (1), 137–150.

Herwartz, H., Siedenburg, F., Walle, Y. M., 2016. Heteroskedasticity robust panel unit root testing under variance breaks in pooled regressions. Econometric Reviews 35 (5), 727–750.

Homm, U., Breitung, J., 2012. Testing for speculative bubbles in stock markets: A comparison of alternative methods. Journal of Financial Econometrics 10 (1), 198–231.

Hsu, Y.-C., Lee, C.-C., Lee, C.-C., 2008. Revisited: Are shocks to energy consumption permanent or temporary? New evidence from a panel SURADF approach. Energy Economics 30 (5), 2314–2330.

Im, K. S., Pesaran, H. M., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115 (1), 53–74.

Joyeux, R., Ripple, R. D., 2007. Household energy consumption versus income and relative standard of living: A panel approach. Energy Policy 35 (1), 50–60.

Kelejian, H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International economic review 40 (2), 509–533.

Kim, T. H., Leybourne, S., Newbold, P., 2002. Unit root tests with a break in innovation variance. Journal of Econometrics 109 (2), 365–387.

Levin, A., Lin, C. F., Chu, C. J., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. Journal of Econometrics 108 (1), 1–24.

McAvinchey, I. D., Yannopoulos, A., 2003. Stationarity, structural change and specification in a demand system: the case of energy. Energy Economics 25 (1), 65–92.

Moon, H. R., Perron, B., 2004. Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122 (1), 81–126.

Narayan, P. K., Smyth, R., 2007. Are shocks to energy consumption permanent or temporary? Evidence from 182 countries. Energy Policy 35 (1), 333–341.

Pesaran, M. H., 2007. A simple panel unit root test in the presence of cross section dependence. Journal of Applied Econometrics 22 (2), 265–312.

Sensier, M., van Dijk, D., 2004. Testing for volatility changes in u.s. macroeconomic time series. Review of Economics and Statistics 86 (3), 833–839.

Stock, J. H., Watson, M. W., 2003. Has the business cycle changed and why? NBER Macroeconomic Annual 17, 159–230.

Westerlund, J., 2014. Heteroscedasticity robust panel unit root tests. Journal of Business & Economic Statistics 32 (1), 112–135.

Westerlund, J., 2015. The effect of recursive detrending on panel unit root tests. Journal of Econometrics 185 (2), 453–467.

B Nonparametric tests for independence – a review and comparative simulation study with an application to malnutrition data in India

Nonparametric Tests for Independence – A Review and Comparative Simulation Study with an Application to Malnutrition Data in India

Helmut Herwartz Simone Maxand*
University of Goettingen

July 10, 2017

Abstract

The detection of dependence structures within a set of random variables provides a valuable basis for a detailed subsequent investigation of their relationships. Beyond common diagnostics for linear correlation, nonparametric tests for independence require only basic assumptions on the marginal or joint distribution of the involved variables. In this paper, we review nonparametric tests of independence in bivariate as well as multivariate settings which are throughout ready-to-use, i.e., implemented in R packages. Highlighting their distinct empirical size and power properties in various small sample settings, our analysis supports an analyst in deciding for a particular test to use with regard to representative underlying distributional settings. Avoiding restrictive moment conditions, the copula based Cramér-von Mises distance of Genest & Rémillard (2004) is remarkably robust under the null hypothesis and powerful under diverse settings that are in line with the alternative hypothesis. Based on distinguished test outcomes in small samples, we detect nonlinear dependence structures between childhood malnutrition indices and possible determinants in an empirical application for India.

Keywords: Tests for independence; nonparametric methods; multivariate independence; spatial ranks; empirical copula; distance covariance.

JEL Classification: C12, C14, C88, I12

*Address correspondence to: Simone Maxand, Chair of Econometrics, University of Goettingen, Humboldtallee 3, D-37073 Goettingen, Germany. Telephone: +49 551 3924604, fax: +49 551 397279, e-mail: *smaxand@uni-goettingen.de*.

1 Introduction

Statistical analyses mostly target at the identification and quantification of dependence structures between the variables of interest. Yet, dependence between random variables can be present in various (e.g., linear or nonlinear) forms. Most commonly, analysts apply standard linear regression models presuming linear dependence structures. Whereas classical procedures, such as Pearson's correlation coefficient (e.g. Pearson, 1920) or Wilks' test (Wilks, 1935), diagnose linear dependence in a parametric framework, they might fail to detect nonlinear and nonmonotone dependence structures. Therefore, nonparametric tests aim at keeping prior assumptions on the variables' distribution under the null hypothesis and their relation under the alternative hypothesis at a minimum.

Classical nonparametric approaches have been developed to test for monotone, but not necessarily linear, bivariate dependence structures by means of ranks. Popular representatives for rank based dependence measures are Kendall's tau (Kendall, 1938) and Spearman's rho (Spearman, 1904). Such bivariate dependence tests, however, might lack consistency under several dependence structures (see Rémillard, 2014, for an example). Against this background, various tests for independence have been developed more recently. These tests are supposed to provide powerful tools to detect various forms of dependence especially if more than two random variables are considered.

As noticed by Josse & Holmes (2014), several test procedures are concurrently employed in distinct research communities. Suggestions of new tests are typically accompanied with comparative evidence gathered from stylized Monte Carlo experiments which use specific types of data (either under the null hypothesis or with regard to particular alternatives). For instance, Josse & Holmes (2014) compare a nonparametric approach based on distance covariances with a multivariate extension of Pearson's correlation coefficient for linear dependence, and Siqueira Santos et al. (2013) compare nonparametric tests with a focus on nonlinear dependence structures typically present in the gene expression literature. Noticing that such comparisons might miss important characteristics of various independence diagnostics under diverse frameworks of data generation, we provide a comprehensive overview on the diversity of nonparametric tests suggested in the recent literature. With particular attention on those procedures that are applicable in multivariate samples, we categorize the tests in regard to their underlying theoretical framework, and distinguish multivariate approaches based on spatial ranks, empirical copulas and distance covariances. Along these lines, we consider representative tests which are examined in more detail. Studying simplified, though representative scenarios for the generation of bivariate and trivariate samples allows to trace the test performances (in finite samples) back to essential characteristics of the data. Since alternative nonparametric tests rely on distinct measures of dependence, our work (i) highlights the signaling content of rival dependence diagnostics under diverse dependence patterns, and (ii) points to the scope of alternative tests under more complex data structures.

In an application to data of childhood malnutrition in India we further illustrate the performance of the tests. We consider a standard regression scenario targeting the explanatory content of several variables on one (resp. two) outcome variables. Specifically, we examine the influence of certain characteristics of the child and it's mother on childhood malnutrition. By means of nonparametric independence tests we diagnose the dependence between malnutrition indices in a bivariate setting, and consider dependence between the bivariate malnutrition index and potential determinants by means of tests of groupwise (in)dependence. The nonparametric framework can be exploited to identify nonlinear and nonmonotone dependence structures as a cornerstone for further analysis of the explicit relation between child malnutrition and its possible determinants.

In the next Section we describe distinguished dependence structures which might be present within a set of p random variables. In Section 3, we briefly characterize the considered test procedures along with some extensions and describe their theoretical background. Section 4 provides the simulation results, followed by the empirical example in Section 5. Section 6 concludes.

Throughout we use the following notation: Univariate continuous real valued random variables are denoted by $x_1, \dots, x_p \in \mathbb{R}$. A set of these random variables of size p_1 and p_2 is denoted by $\boldsymbol{x}_1 = (x_1, \dots, x_{p_1}) \in \mathbb{R}^{p_1}$ and $\boldsymbol{x}_2 = (x_1, \dots, x_{p_2}) \in \mathbb{R}^{p_2}$, respectively. The associated marginal distribution functions are F_{x_k} for $k = 1, \dots, p$, and $F_{\boldsymbol{x}_1}, F_{\boldsymbol{x}_2}$. Furthermore, the joint distribution functions are $F_{x_1,\dots,x_p}$ (for the first two variables F_{x_1,x_2}) and $F_{\boldsymbol{x}_1,\boldsymbol{x}_2}$, respectively. Sample observations are indexed with $i = 1, 2, \dots, n$, such that n is the sample size. A random sample of, for instance, variable x_1 is $\{x_{1,1}, \dots, x_{1,n}\}$. Furthermore, a random sample of the set of variables $\boldsymbol{x}_1$ consists of observations $\boldsymbol{x}_{1,i} = (x_{11,i}, \dots, x_{1p_1,i})'$ for $i \in \{1, \dots, n\}$. The rank of observation $1 \leq i \leq n$ of variable x_k, $1 \leq k \leq p$, is denoted as $R_i^{x_k}$.

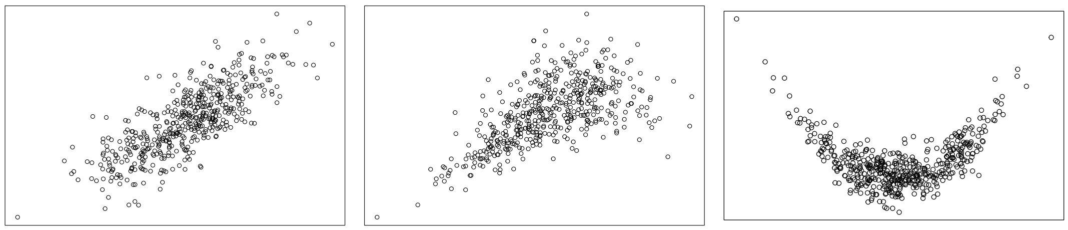


Figure 1: Bivariate standard normal distribution with $\rho = 0.8$ (left), normally distributed variables with Clayton copula with parameter $\theta = 1.5$ (middle) and the functional relationship $x_2 = x_1^2 + \varepsilon$ for $x_1 \sim \mathcal{N}(0, 0.5)$ and $\varepsilon \sim \mathcal{N}(0, 0.2)$ (right).

2 Dependence structures

Quantifying the relation between random variables often relies on the a-priori suggestion of a linear association (see, e.g., the linear positive linkage displayed in the left hand side panel of Figure 1). However, dependence between the variables can not only be characterized by a linear but by means of diverse functional forms. Besides the linear relationship two further examples of dependence structures between two random variables x_1 and x_2 are displayed in Figure 1. The second structure is characterized by dependence in the lower tail of the distributions. Such types of nonlinear dependence are commonly described by means of copulas, i.e. a function C which combines the two marginal distribution functions F_{x_1} and F_{x_2} to the joint distribution function $F_{x_1,x_2}(x_1, x_2) = C(F_{x_1}(x_1), F_{x_2}(x_2))$.[1] Furthermore, a functional nonlinear and nonmonotone association characterizes the relationship between the variables in the right hand side panel of Figure 1.

Although nonparametric tests of the null hypothesis of independence aim at performing adequately irrespective of the underlying distribution, they build upon certain (test specific) regularity assumptions. These might imply performance differences conditional on both the marginal distributions under the null hypothesis and the type of dependence under the alternative hypothesis. Starting from the examples of Figure 1, one might distinguish diverse nonmonotone and nonlinear dependence structures generated by copulas or based on functional associations. Additionally, for specific applications, e.g., economic data, modifications

[1]In Figure 1 we model one sided tail dependencies by means of the Clayton copula. In general, the copula C can be uniquely determined following Sklar's theorem (Sklar, 1959). For a detailed description of dependence modelling by means of copulas see, e.g., Joe (1997).

of these structures might be of interest.[2] We consider several forms of dependence in subsets of a set $\{x_1, \dots, x_p\}$ of univariate random variables $x_1, \dots, x_p \in \mathbb{R}$. Besides pairwise (i.e., bivariate) dependencies the structures can become increasingly complicated in larger sets of random variables with $p > 2$. Next, we outline the null hypotheses of bivariate, groupwise and joint independence.

1. **Bivariate dependence:** As illustrated in Figure 1, with $p = 2$, the considered test procedures assess dependence between two random variables x_1 and x_2. The corresponding null hypothesis is $H_0 : F_{x_1,x_2}(x_1, x_2) = F_{x_1}(x_1)F_{x_2}(x_2)$ with joint distribution function F_{x_1,x_2} and marginals F_{x_1}, F_{x_2}.

2. **Groupwise dependence:** Analyzing two sets of variables can be thought of as a generalization of bivariate dependence tests where two disjoint subsets of $\{x_1, \dots, x_p\}$ are subjected to testing, i.e., $\boldsymbol{x}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{x}_2 \in \mathbb{R}^{p_2}$ such that $p_1 + p_2 = p$. The corresponding null hypothesis is $H_0 : F_{\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{x}_1, \boldsymbol{x}_2) = F_{\boldsymbol{x}_1}(\boldsymbol{x}_1)F_{\boldsymbol{x}_2}(\boldsymbol{x}_2)$ for multivariate distribution functions $F_{\boldsymbol{x}_1,\boldsymbol{x}_2}$, $F_{\boldsymbol{x}_1}$ and $F_{\boldsymbol{x}_2}$. Furthermore, some tests allow to diagnose the dependence between more than two disjoint subsets, where $p_1 + \ldots + p_c = p$ and $c > 2$.

3. **Mutual dependence:** To test for overall independence within a set of random variables $\{x_1, \dots, x_p\}$ the null hypothesis is formulated as $H_0 : F_{x_1,\dots,x_p}(x_1, \dots, x_p) = F_{x_1}(x_1) \cdots F_{x_p}(x_p)$. The tests exploit the fact that mutual independence is equivalent to independence within all subsets of $\{x_1, \dots, x_p\}$. This hypothesis is equivalent to stating groupwise independence and choosing subsets of size $p_1 = p_2 = \ldots = p_c = 1$.

In spite of assessing the same null hypothesis, the considered nonparametric independence tests differ in their theoretical derivation. To identify sources of performance differences, we review the theoretical background of the test procedures in the next section and consider their performance under specific marginal distributions and dependence structures by means of a simulation study in Section 4.

[2]Tests for serial dependence in time series are not explicitly considered here. An overview of corresponding approaches is given in Diks (2009).

3 Tests for independence

Independence diagnostics might be classified into four distinct categories according to their theoretical background. Recently, copula, spatial rank and kernel-based methods have been developed to test nonparametrically for independence in a multivariate framework. For benchmarking purposes we compare these approaches with classical test procedures, namely Hoeffding's D and diagnostics going back to Wilks (1935) in bivariate and multivariate designs, respectively. Throughout this section, we describe the framework of the tests, the test statistics and their empirical formulation. The R packages and functions providing respective implementations are listed in Table 1 at the end of this section.

3.1 Classical tests for independence

The category of classical independence tests consists of widely applied approaches that are frequently implemented in statistical software. Several nonparametric tests for bivariate dependence and one parametric test for multivariate dependence are shortly described in the following.

Pearson's correlation coefficient (e.g., Pearson, 1920) was one of the first measures of linear correlation between two random variables. Shortly after, rank correlation methods as Kendall's tau (Kendall, 1938) and Spearman's rho (Spearman, 1904) were developed to test nonparametrically for independence in bivariate settings. While multivariate extensions of these rank–based statistics are studied in the next section, we consider the nonparametric procedure introduced in Hoeffding (1948) first. This test was further extended by Blum et al. (1961) who tabulate the distribution of Hoeffding's D under the null hypothesis of independence. For two random variables x_1 and x_2, Hoeffding's D builds on the theoretical statistic

$$\Delta_{x_1,x_2} = \int \left[F_{x_1,x_2} - F_{x_1}F_{x_2}\right]^2 dF_{x_1,x_2}, \tag{1}$$

which measures the distance between the joint distribution and the product of marginal distributions in a Cramér-von Mises (CvM) sense. For two random samples of size n, $x_{1,1},\ldots,x_{1,n}$ and $x_{2,1},\ldots,x_{2,n}$ the empirical counterpart of Δ_{x_1,x_2} reads as

$$\mathcal{T}_d = D = \frac{\alpha - 2(n-2)\beta + (n-2)(n-3)\gamma}{n(n-1)(n-2)(n-3)(n-4)}, \tag{2}$$

where

$$\alpha = \sum_{i=1}^{n} (R_i^{x_1} - 1)(R_i^{x_1} - 2)(R_i^{x_2} - 1)(R_i^{x_2} - 2), \tag{3}$$

$$\beta = \sum_{i=1}^{n} (R_i^{x_1} - 2)(R_i^{x_2} - 2)Q_i \quad \text{and} \quad \gamma = \sum_{i=1}^{n} Q_i(Q_i - 1). \tag{4}$$

Here, $R_i^{x_1}$ and $R_i^{x_2}$ are the ranks of observations $x_{1,i}$ and $x_{2,i}$, respectively. Furthermore, Q_i denotes the number of observation pairs $(x_{1,j}, x_{2,j})$ for which the ranks of $x_{1,j}$ and $x_{2,j}$ are both smaller than the ranks of $x_{1,i}$ and $x_{2,i}$, respectively, i.e. $Q_i = \sum_{j=1}^{n} \mathbb{I}\{R_j^{x_1} < R_i^{x_1}\}\mathbb{I}\{R_j^{x_2} < R_i^{x_2}\}$.

The statistic in (2) evaluates dependence between two univariate random variables. Wilks' test (Wilks, 1935) can serve as a benchmark diagnostic in a multivariate set of random variables under the assumption of Gaussianity. For p variables $x_1, \ldots, x_p$, mutual dependence is assessed by means of Wilks' Lambda, i.e.,

$$\mathcal{T}_{Lm} = \lambda_{\text{mut}} = -n \cdot \log \left(\frac{\det\left(\text{cov}(x_1, \ldots, x_p)\right)}{\text{var}(x_1) \cdot \ldots \cdot \text{var}(x_p)} \right). \tag{5}$$

The covariance and the variances in (5) are estimated on the basis of a random sample of $x_1, \ldots, x_p$. Similarly, the statistic

$$\mathcal{T}_{Lg} = \lambda_{\text{groupw}} = -n \cdot \log \left(\frac{\det\left(\text{cov}(x_1, \ldots, x_p)\right)}{\det(\text{cov}(\boldsymbol{x}_1)) \cdot \det(\text{cov}(\boldsymbol{x}_2))} \right) \tag{6}$$

is suitable to test for independence between two groups of variables $\boldsymbol{x}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{x}_2 \in \mathbb{R}^{p_2}$. The empirical versions of the test statistics in (5) and (6) are asymptotically χ^2-distributed with p and 2 degrees of freedom, respectively.

3.2 Tests based on spatial signs and spatial ranks

In the following, we consider two nonparametric analogs to Wilks' test in (6) based on standardized spatial signs and ranks. These dependence measures were introduced in Taskinen et al. (2005) and extend the method of Puri & Sen (1971). More precisely, Kendall's tau and Spearman's rho are formulated in the multivariate setting by means of spatial signs and ranks. Two sets of random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are assumed to follow an elliptically symmetric marginal distribution. Accordingly, the multivariate marginal densities of $\boldsymbol{x}_k$, $k = 1, 2$, can be given as

$$f_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = \det(\Sigma_k)^{-1/2} \exp\left(-\Psi(||\Sigma_k^{-1/2}(\boldsymbol{x}_k - \boldsymbol{\mu}_k)||) \right) \tag{7}$$

for some function $\Psi(\cdot)$, shape matrix Σ_k and location vector $\boldsymbol{\mu}_k$.[3] Furthermore, let $\boldsymbol{z}_{1,i}$ denote a vector of standardized data points of observation i, i.e. $\boldsymbol{z}_{1,i} = \widehat{V}_1^{-1/2}(\boldsymbol{x}_{1,i} - \widehat{\boldsymbol{\mu}}_1)$ with $\widehat{\boldsymbol{\mu}}_1$ being an affine-equivariant location estimator and $\widehat{V}_1$ denoting an estimator of the shape matrix.

Then, for the standardized data points $\boldsymbol{z}_{1,i}$ and $\boldsymbol{z}_{1,j}$, $i, j = 1, \ldots, n$, the vector of *standardized spatial signs* reads as

$$\widehat{\boldsymbol{S}}_{ij}^{(1)} = \begin{cases} \frac{\boldsymbol{z}_{1,i} - \boldsymbol{z}_{1,j}}{((\boldsymbol{z}_{1,i} - \boldsymbol{z}_{1,j})'(\boldsymbol{z}_{1,i} - \boldsymbol{z}_{1,j}))^{1/2}} & \text{if } \boldsymbol{z}_{1,i} - \boldsymbol{z}_{1,j} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Analogously, the standardized spatial sign vector of observations of the second set $\boldsymbol{x}_{2,i} = (x_{21,i}, \ldots, x_{2p_2,i})'$ is defined by $\widehat{\boldsymbol{S}}_{ij}^{(2)}$. The vector of the *standardized spatial ranks* of observation i then results as the average of these signs: $\widehat{\boldsymbol{R}}_i^{x_\bullet} = \frac{1}{n}\sum_{j=1}^{n} \widehat{\boldsymbol{S}}_{ij}^{(\bullet)}$, where $\bullet = 1, 2$.

Using these definitions the multivariate extensions of Kendall's tau and Spearman's rho are

$$\tau^2 = \Big|\Big|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \widehat{\boldsymbol{S}}_{ij}^{(1)} \widehat{\boldsymbol{S}}_{ij}^{(2)\prime}\Big|\Big|^2 \qquad \text{and} \qquad \rho^2 = \Big|\Big|\frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{R}}_i^{x_1} \widehat{\boldsymbol{R}}_i^{x_2\prime}\Big|\Big|^2, \tag{9}$$

respectively, and the corresponding test statistics read as

$$\mathcal{T}_s = \frac{np_1p_2}{4c_1^2c_2^2}\tau^2 \qquad \text{and} \qquad \mathcal{T}_{sr} = \frac{np_1p_2}{c_1^2c_2^2}\rho^2. \tag{10}$$

The constants c_1 and c_2 in (10) depend on the marginals $F_{\boldsymbol{x}_1}$ and $F_{\boldsymbol{x}_2}$. Respective estimates are

$$\hat{c}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{R}}_i^{x_1\prime}\widehat{\boldsymbol{R}}_i^{x_1}\right) \qquad \text{and} \qquad \hat{c}_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{R}}_i^{x_2\prime}\widehat{\boldsymbol{R}}_i^{x_2}\right). \tag{11}$$

The test statistics in (10) are $\chi^2(p_1p_2)$-distributed under the null hypothesis of no dependence between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Furthermore, the tests are efficient for alternatives that are contiguous to an elliptical null distribution. Under these alternatives the test statistics in (10) follow a noncentral χ^2-distribution with noncentrality parameter depending on c_1 and c_2 (Taskinen et al., 2005).

[3] For common choices of Ψ the density $f_{\boldsymbol{x}_k}$ corresponds to the multivariate normal distribution, t-distribution or power exponential function. The scatter matrix Σ_k is a positive definite, symmetric and affine invariant matrix. The metric $||\cdot||$ is any permutation and sign change invariant metric. More detailed descriptions of possible distributions are given in Oja (2010).

3.3 Tests based on the empirical copula

By means of copulas the null hypothesis of mutual independence within a set of random variables $\{x_1, \ldots, x_p\}$ is $H_0 : C(F_{x_1}(x_1), \ldots, F_{x_p}(x_p)) = F_{x_1}(x_1) \cdot \ldots \cdot F_{x_p}(x_p)$, where the function C refers to the corresponding unique copula (Sklar, 1959).

The test procedure considered in the following was introduced in Genest & Rémillard (2004), and further analyzed and extended in subsequent works by Genest et al. (2006), Genest et al. (2007) and Kojadinovic & Holmes (2009). The test statistic is formulated as a Cramér-von Mises (CvM) distance and moreover, applies the decomposition techniques for empirical copulas introduced in Deheuvels (1981). In a first step, a set $\{x_1, \ldots, x_p\}$ of univariate random variables $x_1 \in \mathbb{R}, \ldots, x_p \in \mathbb{R}$ is partitioned into all possible decompositions. The global coefficient for mutual dependence in $\{x_1, \ldots, x_p\}$ then consists of the dependence measures within all decompositions. Let $A \subset S_p = \{1, \ldots, p\}$ denote a possible subset of indices. For instance, in the bivariate case only one single subset $A = \{1, 2\}$ has to be considered.

For subsets of indices A, $B \subset S_p$, the joint copula of $x_1, \ldots, x_p$ is expressed by means of a Möbius decomposition $\mathcal{M}$ which decomposes the copula C as

$$\mathcal{M}_A(C) \equiv \sum_{B \subset A} (-1)^{|A \setminus B|} C(u^B) \prod_{k \in A \setminus B} u_k \tag{12}$$

for $u_1, \ldots, u_p \in [0, 1]$ and $u^B \in [0, 1]^p$ such that

$$u_k^B = \begin{cases} u_k & \text{if } k \in B \\ 1 & \text{if } k \notin B. \end{cases}$$

Mutual independence, i.e. the independence copula, is characterized by the copula C for which $\mathcal{M}_A(C) \equiv 0$ for all $A \subset S_p$.

To test for independence based on a sample of observations, the empirical version of the decomposition in (12) reads as

$$\mathcal{M}_A(\mathbb{C}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \prod_{k \in A} \left[\mathbb{I}\{R_i^{x_k} \leq (n+1) u_k\} - U_n(u_k) \right], \tag{13}$$

where $\mathbb{C}_n$ corresponds to the empirical copula, $R_i^{x_k}$ is the rank of $x_{k,i}$ and U_n is the distribution function of a random variable uniformly distributed on $\{1/(n+1), 2/(n+1), \ldots, n/(n+1)\}$.

The resulting $2^p - p - 1$ CvM statistics (one for each possible decomposition of $A \subset \{1, \dots, p\}$) consist of the decomposition in (13)[4],

$$T_A = \int_{[0,1]^p} \{\mathcal{M}_A(\mathbb{C}_n)\}^2 du, \tag{14}$$

which is calculated as

$$T_A = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \prod_{k \in A} \left[\frac{2n+1}{6n} + \frac{R_i^{x_k}(R_i^{x_k} - 1)}{2n(n+1)} + \frac{R_j^{x_k}(R_j^{x_k} - 1)}{2n(n+1)} - \frac{\max(R_i^{x_k}, R_j^{x_k})}{n+1} \right]. \tag{15}$$

Genest & Rémillard (2004) discuss several methods to obtain a global test statistic for mutual independence in $\{x_1, \dots, x_p\}$. On the one hand, various combination methods of the p-values of T_A are considered. For instance, the Fisher combination of p-values is defined as

$$\mathcal{T}_W = -2 \sum_{|A|>1} \log(p_{T_A}), \tag{16}$$

where p_{T_A} is the p-value of T_A. As an alternative, a measure of mutual dependence can be defined by means of a global CvM functional

$$\mathcal{T}_B = \int_{(0,1)^d} \left\{ \sqrt{n} \left(C_n(u) - \prod_{k=1}^{p} U_n(u_k) \right) \right\} du, \tag{17}$$

with cumulative distribution function U_n of a uniformly distributed variable on $\{1/n, \dots, n/n\}$ and the empirical copula C_n. It is worth mentioning that the combination of p-values has been shown to yield more powerful test procedures than measuring overall dependence based on the test statistic in (17) (Genest & Rémillard, 2004).

The described procedures apply to test for bivariate independence or mutual independence within a set of univariate random variables. Additionally, these tests can be extended in a distribution-free manner to the multivariate case by means of a bootstrap procedure. Kojadinovic & Holmes (2009) derive the bootstrap version of the test for mutual independence between vectors of random variables. Furthermore, Beran et al. (2007) use a similar approach by applying the theory of so-called half-spaces and a CvM statistic to diagnose dependence between random vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

[4]Note that the CvM statistic in (14) forms a multivariate measure of dependence similar to Hoeffding's D in the bivariate case. The expression in (13) defines the distance between the empirical copula, instead of the bivariate empirical joint distribution function, and the distribution under independence. Mutual dependence in a subset $A \subset S_p$ is then measured by combining these distances in the CvM statistic in (14).

3.4 Tests based on distance covariance and kernel-based distances

Whereas the two tests in (10) and the one in (17) extend Kendall's tau, Spearsman's rho and Hoeffding's D, respectively, the dependence coefficient proposed in Székely et al. (2007) processes interpoint distances. For sets of random variables $\boldsymbol{x}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{x}_2 \in \mathbb{R}^{p_2}$ with finite moments, let $\varphi_{x_1}, \varphi_{x_2}$ and φ_{x_1,x_2} denote the marginal and joint characteristic functions, respectively. Székely et al. (2007) introduce the test for independence between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in two versions: On the one hand, based on the *distance covariance*

$$\mathcal{V}^2(\boldsymbol{x}_1, \boldsymbol{x}_2) = ||\varphi_{x_1,x_2}(t,s) - \varphi_{x_1}(t)\varphi_{x_2}(s)||_2^2 \geq 0, \tag{18}$$

and alternatively, based on the *distance correlation*

$$\mathcal{R}^2(\boldsymbol{x}_1, \boldsymbol{x}_2) = \begin{cases} \frac{\mathcal{V}^2(\boldsymbol{x}_1,\boldsymbol{x}_2)}{\sqrt{\mathcal{V}^2(\boldsymbol{x}_1,\boldsymbol{x}_1)\mathcal{V}^2(\boldsymbol{x}_2,\boldsymbol{x}_2)}} & \text{if } \mathcal{V}^2(\boldsymbol{x}_1,\boldsymbol{x}_1)\mathcal{V}^2(\boldsymbol{x}_2,\boldsymbol{x}_2) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

In (18), $||\cdot||_2$ corresponds to the norm in the (weighted) L_2-space of functions on $\mathbb{R}^{p_1+p_2}$. Hence, the dependence measures $\mathcal{V}^2$ and $\mathcal{R}^2$ are zero if and only if the two considered sets $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent. For two random samples, consisting of the vectors $\boldsymbol{x}_{1,i} = (x_{11,i}, \ldots, x_{1p_1,i})'$ and $\boldsymbol{x}_{2,i} = (x_{21,i}, \ldots, x_{2p_2,i})'$, $i = 1, \ldots, n$, the corresponding test statistics are calculated from the sample covariances

$$\begin{aligned} &\mathcal{T}_{dCov} = \mathcal{V}_n^2(\boldsymbol{x}_1, \boldsymbol{x}_2) = \\ &\frac{1}{n^2}\sum_{i,j=1}^{n}\left(|\boldsymbol{x}_{1,i}-\boldsymbol{x}_{1,j}|_{p_1} - \frac{1}{n}\sum_{j=1}^{n}|\boldsymbol{x}_{1,i}-\boldsymbol{x}_{1,j}|_{p_1} - \frac{1}{n}\sum_{j=1}^{n}|\boldsymbol{x}_{1,i}-\boldsymbol{x}_{1,j}|_{p_1} + \frac{1}{n^2}\sum_{i,j=1}^{n}|\boldsymbol{x}_{1,i}-\boldsymbol{x}_{1,j}|_{p_1}\right) \\ &\quad\times\left(|\boldsymbol{x}_{2,i}-\boldsymbol{x}_{2,j}|_{p_2} - \frac{1}{n}\sum_{j=1}^{n}|\boldsymbol{x}_{2,i}-\boldsymbol{x}_{2,j}|_{p_2} - \frac{1}{n}\sum_{j=1}^{n}|\boldsymbol{x}_{2,i}-\boldsymbol{x}_{2,j}|_{p_2} + \frac{1}{n^2}\sum_{i,j=1}^{n}|\boldsymbol{x}_{2,i}-\boldsymbol{x}_{2,j}|_{p_2}\right), \end{aligned} \tag{20}$$

where $|\cdot|_{p_1}$ and $|\cdot|_{p_2}$ denote interpoint Euclidean distances. The empirical version of $\mathcal{R}_n^2$ obtains from inserting $\mathcal{V}_n^2(\boldsymbol{x}_1, \boldsymbol{x}_2)$ into (19). Restricting $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to have finite moments, the test is consistent for any type of dependence. Under the null hypothesis, $n\mathcal{V}_n^2/S \xrightarrow{d} Q$ for $n \to \infty$, where $S = (\frac{1}{n^2}\sum_{i,j=1}^n |\boldsymbol{x}_{1,i} - \boldsymbol{x}_{1,j}|_{p_1})(\frac{1}{n^2}\sum_{i,j=1}^n |\boldsymbol{x}_{2,i} - \boldsymbol{x}_{2,j}|_{p_2})$ and Q is a nonnegative quadratic form of centered Gaussian random variables.

Székely et al. (2007) and Székely & Rizzo (2009) modify and extend the tests based on (18) and (19) in several ways. For instance, the norm used in (18) is generalized to $||\cdot||_\alpha$

which implies a more general weight function and results in α-distance dependence measures. A further specification of the distance covariance is obtained by choosing the covariance with respect to a certain stochastic process. The Brownian motion, for instance, obtains the *Brownian distance covariance.*

Rémillard (2009) argues that the performance of the tests based on (18) and (19) depend on the marginal distributions and further, the statistic in (20) is only applicable to test for independence between two sets of random variables. To address these concerns, Matteson & Tsay (2013) suggest probability integral transformations to avoid the dependence on the marginal distributions. In addition, they provide a test for mutual independence using the fact that the null hypothesis of mutual independence within a set of random variables $\{x_1, \ldots, x_p\}$ is equivalent to $H_0 : \varphi_{x_k,x_{k+}} = \varphi_{x_k}\varphi_{x_{k+}}$ for all $k = 1, \ldots, p-1$ and $k+ = k+1, \ldots, p$.

Sejdinovic et al. (2013) embed the distance covariance within a more general group of dependence measures which has originated from machine learning. The kernel-based so-called Hilbert-Schmidt independence criterion (HSIC) measures the distance between embeddings of distributions into reproducing kernel Hilbert spaces (RKHS). Choosing specific distance induced kernels, the distance covariance then is equivalent to the HSIC based on the RKHS. By linking these two classes of statistics $\mathcal{T}_{dCov}$ might be considered as a representative for HSICs.[5]

4 Performance under specific dependence structures

Although all considered tests have been proposed to evaluate the null hypothesis of independence nonparametrically, their underlying distributional assumptions are more or less restrictive. Especially in small samples this might lead to size and power differentials under the null hypothesis and certain dependence alternatives, respectively. The following simulation study is supposed to identify such performance differentials. We describe the simulation design first and discuss the results afterwards.

[5]We only consider $\mathcal{T}_{dCov}$ and refer to Sejdinovic et al. (2013) for performance comparisons of further HSICs with alternative kernel choices.

Table 1: R packages and functions corresponding to the procedures described in Section 3 and applied within the simulation study. The respective indexation for the corresponding test statistic $\mathcal{T}_\bullet$ used throughout the simulation study is given in parentheses.

	classical	spatial rank	empirical copula	distance covariance
R package	*Hmisc* (Harrell, 2015)	*SpatialNP* (Sirkia et al., 2013)	*copula* (Hofert et al., 2015)	*energy* (Rizzo & Szekely, 2014) *steadyICA* (Risk et al., 2015)
R function for distinct dependence levels				
bivariate	`hoeffd` (d)	`sr.indep.test` (sr)	`indepTest` (B)	`indep.test` ($dCov$)
mutual	Wilks' Lambda (L)	Fisher comb. of sr	`indepTest` (W, B)	`permTest` ($dCov$)
groupwise	Wilks' Lambda (L)	`sr.indep.test` (sr)	`multIndepTest` (B)	`indep.test` ($dCov$)

4.1 Simulation setting

As outlined in Section 2, the considered tests diagnose dependence between two continuous random variables, two or more vectors of variables, or mutual dependence in a set of more than two variables. Within these settings we compare the size and power of the tests either with respect to the implied correlation ρ or the sample size n. The underlying distributional settings are summarized in Table 2.[6]

4.1.1 Bivariate sets of random variables

Random samples $(x_{1,1}, \ldots, x_{1,n})$ and $(x_{2,1}, \ldots, x_{2,n})$ are generated under the null hypothesis (independence) and under the alternative hypothesis (dependence). Two elliptical copulas and one representative of Archimedean copulas determine the dependence structure alternatively. In addition, we study a direct association by means of a nonlinear and nonmonotonic function with noise. Finally, we investigate robustness of the tests to modifications of these dependence structures.

For correlation levels $\rho = 0, 0.1, \ldots, 0.8$,[7] we generate bivariate sets of random variables,

[6] Note that we consider representative distributions and dependence structures which are supposed to unravel differences and similarities of the tests. For more diverse settings, for instance, alternative choices of sample sizes, copulas and marginals, the results are comparable and omitted for space considerations.

[7] More precisely, for elliptical copulas a correlation matrix $V_{x_1,x_2} = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ determines the dependence struc-

and, hence, focus on the correlation moving from the null hypothesis of independence to larger degrees of dependence. We calculate four test statistics, namely Hoeffding's $\mathcal{T}_d$, the Cramér-von Mises statistic $\mathcal{T}_B$, the multivariate extension of Kendall's tau $\mathcal{T}_{sr}$, and the distance covariance $\mathcal{T}_{dCov}$ (see Table 1). The estimated power of the tests is the share of $R = 1000$ test statistics $\mathcal{T}_\bullet$, $\bullet \in \{d, B, sr, dCov\}$, with p-value below the nominal significance level of $\alpha = 0.05$.[8] We provide the size adjusted power with respect to the empirical level $\hat{\alpha}$, and compare the size and power of the tests for sample sizes $n = 10, 50, 100$.

Dependence modeling by means of copulas: Three distinct marginal distributions and a dependence structure determined by three copulas specify the bivariate distribution structure. Regarding the univariate marginal distributions we choose the standard normal, the exponential and the Cauchy distribution. Monotonic and linear dependence is covered by means of the bivariate Gaussian distribution. Moreover, the Student's t- and Clayton copula allow for tail dependencies and thus, represent nonlinear dependence structures. In order to generate respective random samples of size n, we apply the `R` functions `mvdc` and `rMvdc` from the `R` package *copula* (Hofert et al., 2015).

Functional dependence structure: From a distinct perspective, dependence can be seen as an information structure characterizing the data. Increasing the level of noise in a bivariate set of random variables changes the structure from a deterministic relationship to independence. We relate two random variables x_1 and x_2 directly by means of a function, i.e. $x_2 = f(x_1)$, to allow for nonlinear and nonmonotonic types of dependence. As an example, we consider a quadratic structure $x_2 = x_1^2 + \varepsilon$, where $x_1 \sim \mathcal{N}(0, 0.5)$[9], and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian noise term with increasing standard deviation $\sigma = 0, 0.1, \ldots, 1.5$. The variables x_1 and ε are independently drawn in every Monte Carlo iteration indexed by $r = 1, \ldots, R$. Perfect dependence corresponds to $\sigma = 0$, whereas a low level of association is present if $\sigma = 1.5$. A generated sample with $\sigma = 0.2$ has been shown as an example in the right hand side panel of

ture. Archimedean copulas, e.g. the Clayton copula, are formulated with respect to the correlation coefficient by means of a generator function $\psi(\rho)$, $\rho \in [0, \infty)$. For the explicit definition of the copula as a function of correlation ρ we refer to the documentation of the respective `R` functions (Hofert et al., 2015).

[8] In this study, the considered nominal significance level is $\alpha = 0.05$. With respect to other conventional levels, for instance $\alpha = 0.1$, similar results obtain.

[9] Siqueira Santos et al. (2013) consider alternative choices for the distribution of x_1 as, for instance, equidistant points or the uniform distribution. Additionally, they study further nonmonotonic and nonlinear dependence structures, i.e., alternative choices of the function f.

Figure 1.

Modifications of the dependence structures: In practice, the assumption of a homogenous distribution within the entire sample might be not appropriate for an actual data set. For instance, in economic data a varying dependence structure might be present. Furthermore, not only dependence between the marginals might exist but the marginals themselves might incorporate dependence in their variances (see, for instance, Manner & Reznikova, 2012).[10] Allowing for modifications of the distributional settings we consider varying degrees of dependence first. For this purpose, we generate a bivariate normally distributed sample with two distinct levels of correlation, i.e. $\rho_1 = 0.2$ in the first half and $\rho_2 = 0.4$ in the second half of the sample. As a second modification, we formalize dependence among the marginals as implied by a bivariate GARCH process. More explicitly, we sample data from a so-called Constant Conditional Correlation GARCH(1,1) process (CCC-GARCH(1,1), see Bollerslev, 1990). Accordingly, observations $x_{k,i}$, $i = 1, \ldots, n$, $k = 1, 2$, are drawn as

$$x_{k,i} = h_{k,i}^{1/2} z_{k,i} \quad \text{with} \quad h_{k,i} = a_{k0} + a_{kk} x_{k,i-1}^2 + b_{kk} h_{k,i-1} \quad \text{and} \quad \boldsymbol{z}_i \sim \mathcal{N}(0, P), \tag{21}$$

with Gaussian innovations $z_{i,k}$ and $a_{k0} = 1$, $a_{kk} = b_{kk} = 0.4$. Dependence between the univariate GARCH processes is modeled by an unconditional covariance matrix P with $p_{11} = p_{22} = 1$ and off diagonal elements $p_{12} = p_{21} = \rho = 0.4$. For a more detailed description of the CCC-GARCH sampling we refer to the manual of the `R` package *ccgarch* (Nakatani, 2010).

4.1.2 Multivariate sets of random variables

As described in Section 2, a set of more than two random variables might exhibit groupwise or mutual dependence. To uncover differences and similarities between tests for mutual independence, we consider a most simple framework, i.e. a set of three univariate random variables. Within such sets $\{x_1, x_2, x_3\}$ we formalize the dependence structure under the alternative hypothesis by means of equal correlation ρ in bivariate tuples $\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}$. Accordingly, the correlation matrix of $\{x_1, x_2, x_3\}$ reads as

$$V_{x_1,x_2,x_3} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}. \tag{22}$$

[10] Manner & Reznikova (2012) describe, for instance, how such structures complicate the copula representation of the distribution.

Table 2: Simulation settings.

	dependence structure	power performance wrt parameter
bivariate x_1, x_2	normal, t, Clayton copula with normal, exponential, Cauchy marginals	$n = 10, 50, 100$ (size, size adj power for $\rho = 0.4$) $\rho = 0, 0.1, \ldots, 0.8$ (power)
	$x_2 = x_1^2 + \varepsilon, x_1 \sim \mathcal{N}(0, 0.5)$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$	$\sigma = 0, 0.1, \ldots, 1.5$
	varying dependence, $\rho_1 = 0.2$, $\rho_2 = 0.4$	$n = 20, 50, 100$
	CCC-GARCH(1,1), $\rho = 0.4$	$n = 20, 50, 100$
mutual x_1, x_2, x_3	normal copula with normal and Cauchy marginals, $n = 100$	$\rho = 0, 0.1, \ldots, 0.8$
groupwise $\{x_{11}, x_{12}\}, \{x_{22}\}$	normal copula and marginals, $n = 100$, $\rho_{intra} = 0$ and 0.8	$\rho_{inter} = 0, 0.1, \ldots, 0.8$

Similar to the bivariate case, we consider several marginal distributions and three dimensional copulas with increasing levels of correlation $\rho = 0, 0.1, \ldots, 0.8$. We study the performance of two versions of the copula based procedures, namely the global CvM statistic $\mathcal{T}_B$ and the Fisher combination of p-values in subsamples $\mathcal{T}_W$. Moreover, we consider the mutual version of the distance covariance $\mathcal{T}_{dCov}$, a Fisher combination of p-values of the bivariate $\mathcal{T}_{sr}$ statistics and Wilks' lambda $\mathcal{T}_{Lm}$.

Furthermore, we compare tests for groupwise dependence between two sets of variables $\boldsymbol{x}_1 = \{x_1, x_2\}$ and $\boldsymbol{x}_2 = \{x_3\}$. In (22), we have only considered one single correlation level ρ such that independence implies zero correlation globally (i.e., $\rho = 0$). Borrowing from the simulation study in Kojadinovic & Holmes (2009), the correlation matrix employed to formalize groupwise dependence of $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ reads as

$$V_{\boldsymbol{x}_1, \boldsymbol{x}_2} = \begin{pmatrix} 1 & \rho_{intra} & \rho_{inter} \\ \rho_{intra} & 1 & \rho_{inter} \\ \rho_{inter} & \rho_{inter} & 1 \end{pmatrix}. \tag{23}$$

The null hypothesis of groupwise independence corresponds to absence of inter group correlation ($\rho_{inter} = 0$). Accordingly, we study power properties with respect to increasing inter group correlation, i.e. $\rho_{inter} = 0.1, \ldots, 0.8$. Apart from inter group dependence, ρ_{intra}

in (23) denotes the strength of intra group correlation. Intra group correlation ρ_{intra} might differ from zero even under the null hypothesis of groupwise independence. To account for distinct degrees of intra group dependence in the simulation study, we select two distinct levels of correlation within $\boldsymbol{x}_1$, namely, $\rho_{intra} = 0$ (no correlation) and $\rho_{intra} = 0.8$ (strong correlation).

We compare four test statistics under the null hypothesis and the alternative hypothesis of groupwise dependence: the Cramér-von Mises statistic $\mathcal{T}_B$, the statistic $\mathcal{T}_{sr}$ based on spatial ranks, the distance covariance $\mathcal{T}_{dCov}$ and the parametric test based on Wilks' lambda $\mathcal{T}_{Lg}$.

4.2 Simulation results

In the following discussion, the results for size and size adjusted power provide a baseline comparison of the tests. In the subsequent investigation we consider power properties with respect to increasing correlation for copulas and decreasing noise for functional dependence. Furthermore, we address robustness of the tests under modifications of the stylized dependence structures. In multivariate sets of random variables we study the performance of the considered tests under mutual and groupwise dependence alternatives.[11]

4.2.1 Bivariate sets of random variables

Empirical size and size adjusted power

Table 3 documents the estimated size and the size adjusted power of the test statistics $\mathcal{T}_\bullet$, $\bullet \in \{d, B, sr, dCov\}$ with respect to three distinct copulas for sample sizes $n = 10, 50, 100$ and correlation levels $\rho = 0$ (size, in columns 4-7) and 0.4 (power, in columns 8-11). Under respective regularity conditions the statistics $\mathcal{T}_\bullet$ are supposed to converge to the asymptotic distribution for increasing sample sizes (see Section 3). Consequently, the empirical size $\hat{\alpha}$ converges to the true level $\alpha = 0.05$ under these regularity conditions. Deviations from the true level might reflect, on the one hand, violations of the conditions. On the other hand, they contrast the small sample performance of the tests with asymptotic approximations and,

[11] To avoid that test outcomes are fully driven by the underlying distributional setting, it might be interesting to consider feasible combinations of p-values of the rival tests. Due to dependence of the test statistics, an adequate combination might require a sophisticated theoretical derivation and is, therefore, left for future research.

in particular, are informative on the speed of convergence.

We generate bivariate samples under the null hypothesis by means of the respective copula with zero correlation. Although all samples comprise independently drawn marginals, the independence tests perform differently under distinct choices of copulas and marginals. Over all generated samples, the size distortions of the CvM statistic $\mathcal{T}_B$ appear smaller compared with those of the other test procedures. Furthermore, the empirical size of $\mathcal{T}_B$ changes only slightly with respect to the chosen marginal distributions. In contrast, size distortions of $\mathcal{T}_{sr}$ and $\mathcal{T}_d$ are much larger. The two statistics show oversizing in nearly all considered samples with $\mathcal{T}_d$ showing an empirical level closer to the true significance level in comparison with $\mathcal{T}_{sr}$. The statistic $\mathcal{T}_{dCov}$ holds adequate size properties for a sample generated by means of the Gaussian or the Clayton copula. However, we can observe oversizing of this test for the Student's t-copula in combination with all marginals. Under Cauchy marginals, $\mathcal{T}_{dCov}$ shows an empirical level as large as $\hat{\alpha} = 0.461$.

The power estimates displayed in Table 3 are adjusted with respect to the empirical size of the tests. The size adjustment lowers (increases) the rejection frequencies of oversized (undersized) tests to enable a direct comparison of the power of alternative test procedures. Overall, $\mathcal{T}_d$ shows a slight lead in terms of size adjusted power in most scenarios. Under Gaussian copula dependence, the size adjusted power of all tests converges with a similar rate and almost approaches unity for $n = 100$. As the only exception, the size adjusted power of $\mathcal{T}_{dCov}$ fails to converge within the considered sample sizes under the Gaussian copula with Cauchy marginals. For dependence generated by means of a Student's t-copula, the results are similar to the normal dependence structure. Moreover, the tests (except $\mathcal{T}_{dCov}$ for Cauchy marginals) are consistent under dependence modeled by means of a Clayton copula while showing slower convergence rates as under a Gaussian and Student's t-copula.

Theoretically, the inferior performance of the distance covariance $\mathcal{T}_{dCov}$ for specific marginals is in line with its dependence on the marginal distribution (see Section 3). In particular, the moments of the Cauchy distribution are not finite and thus, the regularity conditions that underlie $\mathcal{T}_{dCov}$ do not hold. Overall, the size distortions in small samples indicate which tests might not be appropriate given the underlying distributional setting. The baseline comparison of size and size adjusted power displays comparable test performances in the standard setting, i.e., the Gaussian copula with normal and exponential marginals. However, under a t-copula

Table 3: Empirical size $\hat{\alpha}$ and size adjusted power for correlation $\rho = 0.4$, $R = 1000$ and $\alpha = 0.05$ with respect to alternative bivariate copulas and marginals. Empirical sizes deviating from the true level by more than 0.014 ($\approx 1.96\sqrt{0.05 \cdot 0.95/1000}$) are marked in bold.

			size				size adjusted power			
copula	marginal	n	d	B	sr	$dCov$	d	B	sr	$dCov$
normal	normal	10	**0.099**	0.052	**0.106**	0.050	0.124	0.183	0.181	0.203
		50	**0.066**	0.050	0.058	0.048	0.743	0.748	0.796	0.740
		100	0.063	0.055	0.063	0.055	0.960	0.957	0.977	0.957
	exp.	10	**0.100**	0.046	**0.111**	0.048	0.550	0.171	0.180	0.179
		50	0.062	0.048	0.057	0.049	0.935	0.730	0.778	0.653
		100	0.058	0.050	0.044	0.043	0.992	0.964	0.977	0.948
	Cauchy	10	**0.115**	0.054	**0.112**	0.047	0.148	0.539	0.120	0.099
		50	**0.065**	0.050	0.061	0.051	0.944	0.747	0.790	0.138
		100	0.055	0.041	0.047	0.041	0.995	0.965	0.975	0.209
Student's t	normal	10	**0.135**	0.058	**0.137**	**0.083**	0.157	0.161	0.143	0.137
		50	**0.074**	0.053	**0.070**	**0.085**	0.650	0.673	0.686	0.624
		100	**0.067**	**0.065**	**0.087**	**0.105**	0.941	0.945	0.958	0.929
	exp.	10	**0.131**	0.054	**0.142**	**0.066**	0.506	0.149	0.128	0.156
		50	0.057	0.047	**0.067**	**0.100**	0.912	0.728	0.718	0.517
		100	**0.069**	0.057	**0.070**	**0.137**	0.993	0.945	0.952	0.824
	Cauchy	10	**0.103**	0.053	**0.131**	**0.117**	0.337	0.168	0.421	0.244
		50	**0.071**	0.054	**0.076**	**0.321**	0.543	0.153	0.137	0.052
		100	**0.074**	**0.068**	**0.088**	**0.461**	0.996	0.951	0.951	0.055
Clayton	normal	10	**0.104**	0.046	**0.111**	0.062	0.102	0.121	0.078	0.119
		50	0.061	0.045	0.052	0.047	0.343	0.369	0.412	0.453
		100	0.063	0.059	0.059	0.062	0.648	0.669	0.663	0.707
	exp.	10	**0.120**	0.058	**0.124**	0.044	0.428	0.071	0.063	0.068
		50	0.050	0.040	0.052	0.046	0.762	0.391	0.409	0.242
		100	0.044	0.048	0.058	0.057	0.915	0.709	0.695	0.409
	Cauchy	10	**0.120**	0.051	**0.118**	0.053	0.410	0.079	0.084	0.089
		50	**0.071**	0.056	0.062	0.051	0.764	0.359	0.410	0.173
		100	**0.065**	0.048	0.061	0.057	0.920	0.666	0.674	0.148

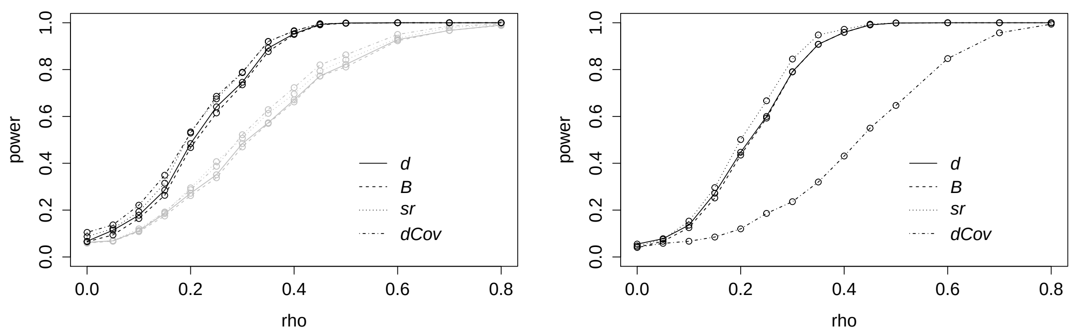


Figure 2: Power curves for bivariate dependence modeled by means of copulas: t-copula (black) and Clayton-copula (grey) with Gaussian marginals (left) and Gaussian copula with Cauchy marginals (right).

we can observe notable differences as described above. Consequently, under tail dependencies the choice of the test appears more crucial for the test decision. In terms of empirical size, the test statistic $\mathcal{T}_B$ seems to perform best irrespective of the underlying distribution. Additionally, we observe that the considered tests show inferior power under dependence governed by the Clayton copula. Moreover, the distance covariance performs weakly for specific marginals.

In order to compare the tests by means of the size adjusted power one has to have in mind that the size adjustments are substantial in case of large size distortions. The size adjusted power of a test serves to compare the test in simulated settings but is, however, not applicable in practice since the empirical level $\hat{\alpha}$ is typically unknown. Therefore, in empirical research one might rather be interested in the comparison of unadjusted power of tests for which it can be safely presumed that their empirical size is close to the nominal level.

Power curves

Figure 2 displays unadjusted power curves with respect to varying levels of correlation $\rho = 0, 0.1, \ldots, 0.8$ and fixed sample size $n = 100$. In particular, for $\rho = 0.4$ the size adjusted counterparts of these empirical power estimates are displayed in Table 3. Studying the power curves, our interest is in their overall shape or, more specifically, in the degree of dependence that obtains a power of unity such that the tests detect dependence with probability one.

First, we contrast the results for two representative copula structures in the left hand side panel of Figure 2, namely the t-copula (black curves) and the Clayton-copula (grey) both combined with Gaussian marginals. The resulting empirical sizes for $\rho = 0$ are comparable with the ones displayed in Table 3. In both settings the power curves of the considered tests have a similar shape. Power of unity is attained for a similar level of correlation for the t-copula ($\rho \approx 0.5$) and for the Clayton-copula ($\rho \approx 0.8$). In this sense, the power of all four tests converges to unity faster in case of dependence generated from the t-copula than for dependence emerging from the Clayton-copula. The procedures are consistent against both alternatives. Overall, $\mathcal{T}_{dCov}$ and $\mathcal{T}_{sr}$ slightly outperform the independence diagnostics based on the CvM statistic in these scenarios.

Furthermore, the right hand side panel of Figure 2 displays power curves for a Gaussian copula with Cauchy marginals. In line with the results shown in Table 3, the power curve of $\mathcal{T}_{dCov}$ stays throughout remarkably below the other curves. Especially, under Cauchy marginals with non existing moments $\mathcal{T}_{dCov}$ suffers from power weakness. In addition, having also in mind the size distortions under the t-copula with Cauchy marginals, $\mathcal{T}_{dCov}$ might not be appropriate under these specific marginal distributions. Nevertheless, for alternatives far away from the null hypothesis of independence ($\rho = 0.8$) all tests show power of unity.

Besides copula dependencies we relate the variables x_1 and x_2 in a functional manner $x_2 = x_1^2 + \varepsilon$ to represent a nonlinear and nonmonotone dependence alternative. Rejection frequencies for samples of size $n = 100$ are depicted in the left hand side panel of Figure 3.[12] Except the spatial rank based procedure $\mathcal{T}_{sr}$ the power of all tests is unity for the deterministic relationship $x_2 = x_1^2$ (i.e. $\sigma = 0$) up to moderate levels of uncertainty ($\sigma = 0.3$). Power estimates for both the CvM statistic $\mathcal{T}_B$ and Hoeffding's $\mathcal{T}_d$ increase with a decreasing level of uncertainty, but are throughout smaller in comparison with the power of $\mathcal{T}_{dCov}$. In contrast, the spatial rank based procedure $\mathcal{T}_{sr}$ indicates the deterministic association $x_2 = x_1^2$ in only 30% of the cases. For the convergence of $\mathcal{T}_{sr}$, Taskinen et al. (2005) assume an elliptical distribution so that the procedure is not necessarily consistent against the nonmonotone alternative. Furthermore, Ding & Li (2014) argue that dependence structures formalized as functional relationships might correspond to singular copulas. A singular copula violates the assumption of absolutely continuous copulas imposed by Genest & Rémillard (2004). Thus, for such a dependence structure $\mathcal{T}_{dCov}$ might be preferred over $\mathcal{T}_d$ and $\mathcal{T}_B$ while $\mathcal{T}_{sr}$ suffers

[12] Starting with the deterministic relationship and modeling dependence up to a certain level of noise, we are interested if this type of dependence is detected, rather than in the test behavior close to/under the null hypothesis.

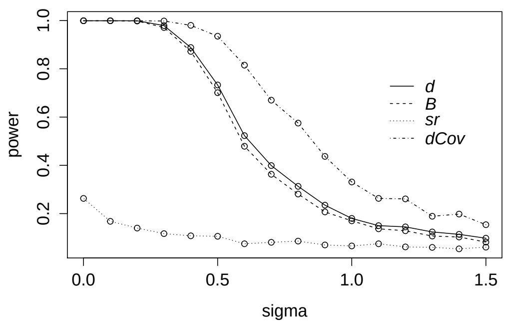


Figure 3: Power curve for quadratic dependence with respect to the standard deviation of the noise term ($\sigma = 0$ refers to perfect dependence and $\sigma = 1.5$ to weak association).

		power			
Type of modification	n	d	B	sr	$dCov$
Varying dependence,	20	0.269	0.210	0.308	0.259
$\rho_1 = 0.2$, $\rho_2 = 0.4$	50	0.520	0.497	0.564	0.568
	100	0.781	0.775	0.829	0.835
CCC-GARCH(1,1)	20	0.363	0.323	0.404	0.342
	50	0.714	0.704	0.743	0.709
	100	0.938	0.942	0.956	0.930

Table 4: Power with respect to different modifications of the dependence structure and marginals for a bivariate normal distribution.

from prohibitive power loss.

In summary, the results for the power curves allow similar conclusions as those documented for empirical size and size adjusted power of the tests. Standard distributional settings lead to comparable performances of all tests. In particular, oversizing under the t-copula and the inferior performance of $\mathcal{T}_{dCov}$ are notable in this respect. Furthermore, one would rank the tests differently based on their performance under a nonmonotone dependence structure.

Robustness to modifications

Results documented in Table 4 address the robustness of the tests to non standard data structures for samples of size $n = 20, 50, 100$. In heterogeneous random samples a varying dependence structure (compared with constant dependence in the entire sample) might be present. The results for a bivariate normal distribution with $\rho_1 = 0.2$ and $\rho_2 = 0.4$ indicate that all considered test procedures remain consistent. The power for $n = 100$ is, in fact, comparable with rejection frequencies in a sample with homogeneous correlation $\rho = 0.3$. For larger samples or stronger levels of correlation all procedures show satisfactory power properties. Nevertheless, the power estimates of $\mathcal{T}_d$ and $\mathcal{T}_B$ converge slower than their counterparts obtained from $\mathcal{T}_{sr}$ and $\mathcal{T}_{dCov}$.

Table 4 documents the results for a normally distributed CCC-GARCH(1,1) process with unconditional correlation of $\rho = 0.5$. As it turns out, all considered tests are consistent against this type of dependence with comparable speed of convergence.[13] In particular, $\mathcal{T}_{sr}$ shows slight power leads in small samples. Overall, the tests are robust under this data structure.

Summarizing the results for the bivariate case, we cannot identify a single nonparametric test which is most powerful against all alternatives. Instead, the size and power performance differs for distinct types of data. We have discovered dependence structures where slight differences between the tests are identifiable, as well as structures where the test decision might depend more strongly on the choice of the test. Based on its empirical size properties, the CvM statistic $\mathcal{T}_B$ might be preferred as it shows the most stable results. Irrespective of distinct dependence structures the empirical level of $\mathcal{T}_B$ is close to the nominal level of $\alpha = 0.05$. The other tests show oversizing especially under a t-copula, and the distance covariance $\mathcal{T}_{dCov}$ performs worst under the considered copula dependence structures. Nevertheless, $\mathcal{T}_{dCov}$ outperforms the other tests under nonmonotone dependence structures where, in contrast, $\mathcal{T}_{sr}$ shows severe lacks of power.

4.2.2 Multivariate sets of random variables

Multivariate nonparametric independence tests are supposed to have power against alternative hypotheses of mutual and groupwise dependence. In the following, we consider mutual dependence first. Being representative for diverse copula structures the power curves under a Gaussian copula with Gaussian and Cauchy marginals are displayed in Figure 4 for increasing levels of correlation ρ among all pairs of variables.[14] For $\rho = 0$, the considered tests exhibit an empirical level close to $\alpha = 0.05$. Similar to the bivariate scenarios, the shape of all power curves shows comparable characteristics for a multivariate Gaussian distribution. The curves displayed in the left hand side panel of Figure 4 uncover slight power differences between the distinct test procedures for correlation levels between $\rho = 0$ and 0.4. In particular, the CvM distance $\mathcal{T}_B$ appears to outperform the other tests.

As displayed in the right hand side panel of Figure 4, both the distance covariance $\mathcal{T}_{dCov}$ and Wilks' Lambda $\mathcal{T}_{Lm}$ perform poorly under Cauchy marginals in terms of power. The

[13] Power of unity is achieved for $n = 120$.

[14] The asymptotic properties of Wilks' Lambda have been shown under the multivariate Gaussian distribution. Thus, the comparison with the nonparametric tests is informative on the trade-off between efficient dependence detection within the Gaussian model, and robustness under more general distributional conditions.

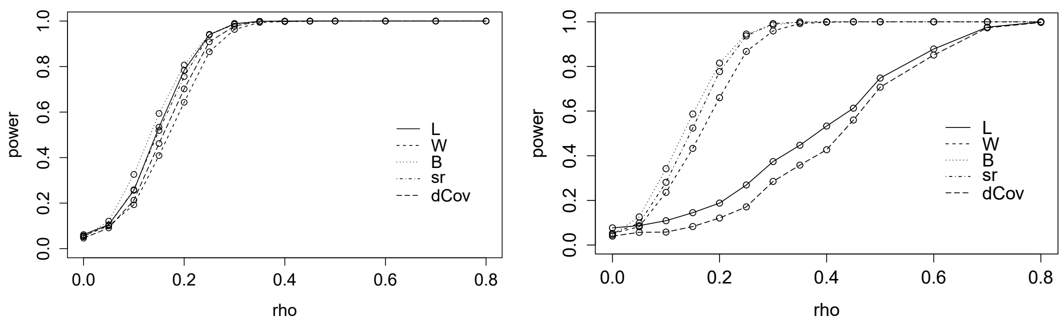


Figure 4: Power curves for mutual dependence as implied by a Gaussian copula with Gaussian (left hand side) and Cauchy (right hand side) marginals: Wilks' Lambda $\mathcal{T}_L$, the Fisher combination of p-values $\mathcal{T}_W$ and $\mathcal{T}_{sr}$, the global CvM statistic $\mathcal{T}_B$ and the distance covariance $\mathcal{T}_{dCov}$.

distance covariance $\mathcal{T}_{dCov}$ might suffer from power losses under a distribution lacking finite moments (cf. Section 4.2.1), while the parametric test $\mathcal{T}_{Lm}$ relies on the assumption of Gaussian distributed variables (see Wilks, 1935). Furthermore, Wilks' Lambda $\mathcal{T}_{Lm}$ exceeds the nominal significance level of $\alpha = 0.05$ under Cauchy marginals.

In summary, in the considered multivariate sets the tests perform in analogy to the bivariate case under distinct marginals and copulas, nonlinear nonmonotone dependence structures and the further modifications. Nevertheless, it is worth mentioning that, e.g., performance differences between $\mathcal{T}_B$ and $\mathcal{T}_W$ might reflect distinct combinations of p-values. Given the results in Genest et al. (2007), heterogeneous power properties (more precisely, power leads of $\mathcal{T}_W$) might result in higher dimensions for which the combination method might become more important.

If the random variables can be aggregated to groups, it might be more interesting to analyze the strength of dependence between the groups of variables (and not within the groups). In Figure 5, power curves are shown for a trivariate set of Gaussian variables with dependence between the marginals determined by means of the covariance matrix in (23). Intra group dependence is fixed whereas inter group dependence varies between 0 and 0.7. The power curves in Figure 5, for $\rho_{intra} = 0$ (black) and $\rho_{intra} = 0.8$ (grey), show characteristics which are comparable with the results for alternatives of mutual dependence. In both cases the power of all test statistics equals unity for levels of inter group correlation in excess of $\rho_{inter} = 0.5$. However, the higher ρ_{intra} the slower is the convergence to a power of unity. The performance differences between the tests are relatively small for the standard

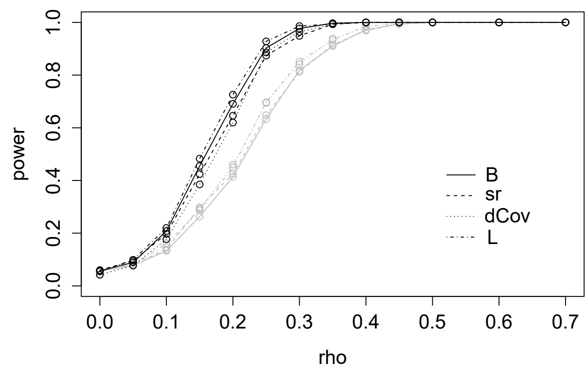


Figure 5: Power curves for multivariate groupwise dependence modeled by means of the Gaussian copula, intra group dependence modeled as in (23) with $\rho_{intra} = 0$ (black) and $\rho_{intra} = 0.8$ (grey).

copula structures. The resulting power properties for variations of marginals and copulas are not displayed here for space considerations but show qualitatively identical characteristics as discussed above for the multivariate dependence structures.[15]

5 Diagnosing dependence patterns for childhood undernutrition

After analyzing diverse pairwise, groupwise and mutual dependence structures by means of a simulation study, this section illustrates the performance of the tests by means of an application to empirical data. We consider data for childhood undernutrition, one of the most urgent public health challenges in developing and transition countries. In studying these data, we are interested in the relation between distinguished measures of undernutrition and a set of child's and mother's characteristics recorded in 1998/99 in the state Uttar Pradesh in Northern India (provided by Demographic and Health Surveys, DHS, www.measuredhs.com). In the following, we apply independence tests to subsamples of $n = 87$ and $n = 55$ children at the age of 3 month ($cage = 3$) and 9 month ($cage = 9$), respectively.

The impact of certain characteristics on undernutrition might be quantified by means of

[15] Kojadinovic & Holmes (2009) and Taskinen et al. (2005), for instance, compare their suggested tests (namely, $\mathcal{T}_B$ and $\mathcal{T}_{sr}$) with Wilks' Lambda $\mathcal{T}_L$ in higher dimensions.

regression models (see, for instance, Kandala et al., 2001; Klein & Kneib, 2016). As a prerequisite, the considered nonparametric independence tests provide guidance for such further analysis. Childhood undernutrition itself can be measured by means of three distinct criteria. First, acute undernutrition (*wasting*) measures insufficient weight for given height. Second, chronic undernutrition (*stunting*) measures insufficient height given age. Third, both forms of undernutrition are captured by means of measuring insufficient weight given age (*underweight*). We apply bivariate and groupwise independence tests to study the relationship between two undernutrition measures, namely *wasting* and *underweight*,[16] and their relationship to two of their possible determinants, namely the mother's age and the mother's body mass index.

First, we apply the bivariate independence tests Wilks' Lambda $\mathcal{T}_L$, the CvM statistic $\mathcal{T}_B$, the spatial rank based statistic $\mathcal{T}_{sr}$ and the distance covariance $\mathcal{T}_{dCov}$ to the set {*underweight*, *wasting*}. The test statistics and the corresponding p-values are displayed in Table 5. The test results indicate dependence between these two indices of malnutrition at a significance level of $\alpha = 0.1$ for both samples with $cage = 3$ and $cage = 9$. Furthermore, all tests except for Wilks' Lambda $\mathcal{T}_L$ indicate significant dependence at level $\alpha = 0.05$ in the sample of $cage = 3$. The distinguished outcomes of $\mathcal{T}_L$ and the nonparametric tests for $cage = 3$ might result from an underlying dependence structure that differs from the bivariate normal model (for instance, including tail dependence). Accordingly, the level of dependence between the indices might be stronger for more extreme levels of undernutrition. Additionally, the dependence between *underweight* and *wasting* is indicated to be stronger in the second sample ($cage = 9$), since the corresponding p-values are throughout below 0.005. For older children ($cage = 9$) it might be more likely that both forms of undernutrition, rather than only one, are observed jointly.

Moreover, we investigate the dependence between the two dimensional set of {*underweight*, *wasting*} and two of the mother's characteristics, namely, the mother's age at birth (*mage*) and the mother's body mass index (*mbmi*).[17] We apply the same tests as in the bivariate setting in their multivariate form (studied in Section 4.1.2) to the two dimensional set of mal-

[16]The results for the remaining bivariate combinations of indices, e.g. {*underweight*, *stunting*}, are comparable to those displayed here.

[17]The whole set of characteristics, i.e. possible covariates in a regression model, are listed in Klein & Kneib (2016) and the references therein. Klein & Kneib (2016) further describe the nonlinear effects of the covariates on the bivariate distribution using nutrition data from all over India.

Table 5: Independence test results for the set $\{underweight, wasting\}$ with respect to the child's age ($cage = 3$ and $cage = 9$ month) and based on samples of sizes $n = 87$ and $n = 55$, respectively.

$\{underweight, wasting\}$		$\mathcal{T}_L$	$\mathcal{T}_B$	$\mathcal{T}_{sr}$	$\mathcal{T}_{dCov}$
$n = 87$, $cage = 3$	statistic	5.163	0.090	7.956	20.946
	p-value	0.076	0.009	0.005	0.01
$n = 55$, $cage = 9$	statistic	22.232	0.219	23.657	40.938
	p-value	0.000	0.001	0.000	0.005

nutrition indices $\boldsymbol{x}_1 = \{underweight, wasting\}$ and one further characteristic being either $\boldsymbol{x}_2 = \{mage\}$ or $\boldsymbol{x}_2 = \{mbmi\}$, respectively.

The test results are documented in Table 6. We can diagnose marked differences between the test outcomes. Studying the dependence between malnutrition and the mother's age, i.e. $\boldsymbol{x}_2 = \{mage\}$, none of the considered tests except for Wilks' Lambda $\mathcal{T}_L$ for $cage = 3$ leads to a rejection of the independence hypothesis with 10% significance. For $cage = 9$ the p-values of the CvM statistic $\mathcal{T}_B$ and the distance covariance $\mathcal{T}_{dCov}$ are smaller but still do not indicate dependence in the second sample ($cage = 9$) with significance of 10%. In contrast, the p-values of Wilks' Lambda $\mathcal{T}_L$ and the spatial rank based statistic $\mathcal{T}_{sr}$ are larger in the sample of nine month old children in comparison with three month old children. In light of the simulation results discussed in Section 4.2 this discrepancy could, on the one hand, reflect a nonlinear relationship that differs from an elliptical distribution. On the other hand, the smaller sample size and the stronger dependence within $\boldsymbol{x}_1$, i.e. between *underweight* and *wasting*, could explain performance weaknesses in the sample with $cage = 9$ (see Section 4.2.2).

In contrast, the test results partly indicate dependence between the mothers's body mass index and the two dimensional undernutrition index of their children. For instance, the null hypothesis of independence is rejected with 10% significance by means of the CvM statistic $\mathcal{T}_B$ and the distance covariance $\mathcal{T}_{dCov}$ in both samples ($cage = 3, 9$). Based on the sign rank based statistic $\mathcal{T}_{sr}$ we can only diagnose dependence for $cage = 9$. In contrast, by means of Wilks' Lambda $\mathcal{T}_L$ independence cannot be rejected and throughout, the p-values are even larger for $cage = 9$ in comparison with $cage = 3$. These distinct test results point to a

Table 6: Independence test results for the sets $\boldsymbol{x}_1 = \{underweight, wasting\}$ and $\boldsymbol{x}_2 = \{mage\}$ or $\boldsymbol{x}_2 = \{mbmi\}$ with respect to the child's age ($cage = 3$ and $cage = 9$ month) and based on samples of sizes $n = 87$ and $n = 55$, respectively.

$\boldsymbol{x}_1$=$\{underweight, wasting\}$			$\mathcal{T}_L$	$\mathcal{T}_B$	$\mathcal{T}_{sr}$	$\mathcal{T}_{dCov}$
$\boldsymbol{x}_2$=$\{mage\}$	$cage = 3$	statistic	5.542	0.015	2.027	3.476
		p-value	0.063	0.840	0.363	0.585
	$cage = 9$	statistic	1.356	0.026	1.527	4.642
		p-value	0.508	0.363	0.466	0.420
$\boldsymbol{x}_2$=$\{mbmi\}$	$cage = 3$	statistic	3.490	0.053	4.150	2.920
		p-value	0.175	0.041	0.126	0.055
	$cage = 9$	statistic	2.620	0.082	6.229	3.568
		p-value	0.270	0.007	0.044	0.070

nonlinear, possibly nonmonotone, and at least non Gaussian dependence structure.

Overall, the test results are in line with the results of Klein & Kneib (2016) who study the dependence between childhood undernutrition and a set of the child's and their mother's characteristics by means of copula regressions for data from all over India. Our results show that the dependence for $cage = 9$ is stronger as it is for $cage = 3$, and might exhibit a non elliptical distribution in both samples. In line with our dependence diagnosis Klein & Kneib (2016) predict the dependence between *wasting* and *underweight* by means of a bivariate Clayton copula obtaining a larger dependence coefficient in the sample of children aged 9 months.

Applying the multivariate tests we have detected dependence between the mother's body mass index (*mbmi*) and the bivariate set of undernutrition measures, and we are led to expect a nonlinear form of dependence. Furthermore, the relation between the mother's age (*mage*) and the undernutrition indices $\{underweight, wasting\}$ lacks significance. Testing for independence between distinct combinations of possible covariates and the bivariate response variable might serve to select covariates with significant explanatory content.[18] In comparison, based on variable selection criteria Klein & Kneib (2016) include *mage* and *mbmi* within their

[18]Note that the tests can further be applied to larger sets of variables to test for groupwise dependence structures. The results are not shown here for space considerations.

regression model to exhibit a nonlinear effect.

In summary, performance differences between the considered tests show up in most of the samples of the nutrition data. The independence tests benefit from satisfactory power even for samples of small size. Moreover, the multivariate tests are applicable to large sets of variables to diagnose between or within dependence in a flexible way. Thereby, dependence between the undernutrition measures and the set of all determining characteristics, as well as mutual dependence within the set of indices could be assessed in further investigations.

6 Conclusions

Nonparametric tests for independence provide a useful basis to decide if the multivariate distribution of random variables merely relies on their marginal distributions, or if it is worth to undertake the specification of a dependence structure. Meeting basic distributional assumptions, nonparametric independence tests have been developed to detect various forms of dependence between two or more random variables. We have described several dependence structures fundamentally, and provided a comprehensive overview of the theoretical background of multivariate nonparametric independence tests. More precisely, the review comprises traditional tests, as well as more recently suggested approaches based on spatial ranks, the empirical copula and the distance covariance.

In a comparative simulation study we consider diverse distributional settings, such as (non)linear copula dependencies, nonmonotone structures and some modifications which point at diverse potential applications. A simulation study unravels distinguished size and power properties under the null hypothesis and specific dependence alternatives, respectively. As a general conclusion, our results do not indicate one overall most powerful test. Rather, the form of dependence appears crucial for the tests to perform preferably. Whereas under multivariate normality the tests show almost equivalent performance, the choice of the tests should be made more cautiously under non Gaussian distributional settings. In particular, the distance covariance performs poorly under distributions which lack finite moments. Furthermore, one might not be able to detect a nonmonotone nonlinear dependence structure by means of spatial rank based tests whereas the distance covariance performs best under this dependence alternative. The test based on the Cramér-von Mises (CvM) statistic seems to be most robust to the diversity of dependence structures. Generally, merits and drawbacks of the alternative tests found in bivariate settings are confirmed for trivariate tests on mutual and groupwise dependence.

In an application to malnutrition data we find that distinguished test outcomes are informative for diverse forms of dependence between the variables and its strength even in samples of small size. Consequently, their nonlinear relation might be subjected to further analysis, for instance, by means of a semiparametric regressions.

The literature on nonparametric independence tests is growing, and already provides refinements of the methods discussed in this work. For instance, Ding & Li (2014) combine the distance covariance and copula based measures which might lead to power gains in the case of a singular copula. Similarly, the set of Hilbert-Schmidt independence criteria (Sejdinovic et al., 2013) promise improvements of dependence diagnosis over the stylized nonparametric approaches compared here. While our results hint at test specific performance patterns, it appears a fruitful avenue of future research to characterize merits and risks of most recent dependence diagnostics under diverse distributional settings and higher dimensionality by means of simulation studies.

References

Bakirov, N. K., Rizzo, M. L., & Székely, G. J. (2006). A multivariate nonparametric test of independence. *J. Multivariate Anal.*, *97*(8), 1742–1756.

Beran, R., Bilodeau, M., & Lafaye de Micheaux, P. (2007). Nonparametric tests of independence between random vectors. *J. Multivariate Anal.*, *98*(9), 1805–1824.

Blomqvist, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statistics*, *21*, 593–600.

Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, *32*, 485–498.

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.

Deheuvels, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.*, *11*(1), 102–113.

Diks, C. (2009). Nonparametric tests for independence. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems* (pp. 6252–6271). Springer New York.

Ding, A. A. & Li, Y. (2014). Copula correlation : An equitable dependence measure and extension of pearson's correlation. arXiv:1312.7214v3.

Fisher, R. (1932). *Statistical methods for research workers*, volume 4. Oliver and Boyd, London.

Genest, C., Quessy, J.-F., & Rémillard, B. (2006). Local efficiency of a Cramér-von Mises test of independence. *J. Multivariate Anal.*, *97*(1), 274–294.

Genest, C., Quessy, J.-F., & Remillard, B. (2007). Asymptotic local efficiency of Cramér-von Mises tests for multivariate independence. *Ann. Statist.*, *35*(1), 166–191.

Genest, C. & Rémillard, B. (2004). Tests of independence and randomness based on the empirical copula process. *Test*, *13*(2), 335–370.

Ghalanos, A. (2015). *rmgarch: Multivariate GARCH models.* R package version 1.2-9.

Harrell, F. E. (2015). *Hmisc*. R package version 3.15-0.

Hoeffding, W. (1948). A non-parametric test of independence. *Ann. Math. Statistics*, *19*, 546–557.

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2015). *copula: Multivariate Dependence with Copulas.* R package version 0.999-13.

Hua, W.-Y., Reiss, P., & Ghosh, D. (2014). Optimal kernel combination for test of independence against local alternatives. arXiv:1409.3636v1.

Ivan Kojadinovic & Jun Yan (2010). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, *34*(9), 1–20.

Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London.

Josse, J. & Holmes, S. (2014). Measures of dependence between random vectors and tests of independence. literature review. arXiv:1307.7383v3.

Jun Yan (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, *21*(4), 1–21.

Kandala, N., Lang, S., Klasen, S., & Fahrmeir, L. (2001). Semiparametric analysis of the socio-demographic and spatial determinants of undernutrition in two african countries. *Research in Official Statistics*, *1*, 81–100.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81–93.

Klein, N. & Kneib, T. (2016). Simultaneous inference in structured additive conditional copula regression models: a unifying bayesian approach. *Stat Comput*, *26*(4), 841–860.

Kojadinovic, I. & Holmes, M. (2009). Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. *J. Multivariate Anal.*, *100*(6), 1137–1154.

Kost, J. T. & McDermott, M. P. (2002). Combining dependent p-values. *Statist. Probab. Lett.*, *60*(2), 183–190.

Manner, H. & Reznikova, O. (2012). A survey on time-varying copulas: Specification, simulations, and application. *Econometric Reviews*, *31*(6), 654–687.

Matteson, D. S. & Tsay, R. S. (2013). Independent component analysis via distance covariance. *pre-print*. http://arxiv.org/abs/1306.4911.

Nakatani, T. (2010). *Conditional Correlation GARCH models.* R package version 0.2.3.

Nelsen, R. B. (1999). *An introduction to copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Oja, H. (2010). *Multivariate nonparametric methods with R*, volume 199 of *Lecture Notes in Statistics*. Springer, New York. An approach based on spatial signs and ranks.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, (13).

Puri, M. L. & Sen, P. K. (1971). *Nonparametric methods in multivariate analysis.* John Wiley & Sons, Inc., New York-London-Sydney.

Rémillard, B. (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Stat.*, *3*(4), 1295–1298.

Rémillard, B. (2014). Tests of independence. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1598–1601). Springer Berlin Heidelberg.

Reshef, Y. A., Reshef, D. N., Sabeti, P. C., & Mitzenmacher, M. (2015). Theoretical foundations of equitability and the maximal information coefficient. arXiv:1408.4908.

Risk, B. B., James, N. A., & Matteson, D. S. (2015). *steadyICA: ICA and Tests of Independence via Multivariate Distance Covariance*. R package version 1.0.

Rizzo, M. L. & Szekely, G. J. (2014). *energy: E-statistics (energy statistics)*. R package version 1.6.2.

Robert, P. & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the *RV*-coefficient. *J. Roy. Statist. Soc. Ser. C Appl. Statist.*, *25*(3), 257–265.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., & Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, *41*(5), 2263–2291.

Siqueira Santos, S. d., Takahashi, D. Y., Nakata, A., & Fujita, A. (2013). A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in bioinformatics*, *15*(6), 906–18.

Sirkia, S., Miettinen, J., Nordhausen, K., Oja, H., & Taskinen, S. (2013). *SpatialNP: Multivariate nonparametric methods based on spatial signs and ranks*. R package version 1.1-1.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, *8*, 229–231.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101.

Székely, G. J. & Rizzo, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.*, *3*(4), 1236–1265.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.*, *35*(6), 2769–2794.

Taskinen, S., Kankainen, A., & Oja, H. (2003a). Sign test of independence between two random vectors. *Statist. Probab. Lett.*, *62*(1), 9–21.

Taskinen, S., Kankainen, A., & Oja, H. (2003b). Tests of independence based on sign and rank covariances. In *Developments in robust statistics (Vorau, 2001)* (pp. 387–403). Physica, Heidelberg.

Taskinen, S., Oja, H., & Randles, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.*, *100*(471), 916–925.

Wilks, S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica*, *3*(3), 309–326.

C Independence based identification of structural shocks: Performance evaluation by means of Monte Carlo simulations and an application to the global crude oil market

Independence Based Identification of Structural Shocks: Performance Evaluation by means of Monte Carlo Simulations and an Application to the Global Crude Oil Market

Helmut Herwartz Simone Maxand*
University of Göttingen

July 10, 2017

Abstract

Structural vector autoregressive analysis aims to trace back the contemporaneous linkages among reduced form disturbances to an interplay of orthogonal structural shocks of unit variance. To identify this interplay the econometric analysis has to rely on additional (often external and not data-based) information. While uncorrelated Gaussian shocks are independent by implication, the often reasonable assumption of non-Gaussian model disturbances offers a new possibility to identify independent structural shocks. We compare the performance of three alternative independence based identification procedures and identification by means of sign restrictions under distinct distributional settings and sample sizes. Thereby, we confirm a bias induced by stylized sign restrictions and find considerable differences between parametric and nonparametric identification schemes that exploit the supposed independence of structural shocks. In an application to the global crude oil market independence based identification performs comparable with techniques of former studies without the need to set up strong economic or distributional assumptions a-priori.

Keywords: Structural shocks, identification, SVAR, global oil market.

JEL Classification: C32, E31, Q43.

*Address correspondence to: Simone Maxand, Chair of Econometrics, University of Goettingen, Humboldtallee 3, D-37073 Göttingen, Germany. Telephone: +49 551 3924604, fax: +49 551 397279, e-mail: *smaxand@uni-goettingen.de*.

1 Introduction

The literature on structural vector autoregressive (SVAR) models provides diverse identification strategies to disentangle the impact of isolated structural shocks. Identification based on economic theory has been widely applied in macroeconometrics. Theory based a-priori assumptions are implemented, for instance, by means of sign restrictions or zero restrictions on the impact or the long-run effects of structural shocks (for instance, Faust, 1998; Sims, 1980; Blanchard and Quah, 1989). Assuming a recursive causation scheme for simplicity, i.e., lower triangularity of the structural matrix, might be too restrictive in many applications. In contrast, well established sign restrictions facilitate the derivation of impulse response functions in that they reduce the set of possible decompositions of the reduced form covariance matrix to those that are in line with a-priori economic reasoning. Nevertheless, the most reasonable decomposition cannot be recovered if its sign pattern does not coincide with the one assumed (Fry and Pagan, 2007).

Against this background, data-driven identification procedures attract specific interest especially if no well-founded theory based assumptions on signs or magnitudes of the effects of structural shocks are available a-priori. Along these lines, introducing additional assumptions on the distributions of the error terms may offer an alternative basis for identification. For instance, Lanne and Lütkepohl (2010) assume mixed normally distributed model disturbances to identify structural shocks and impulse responses. Similar to Rigobon (2003) and Lewbel (2010), Lanne and Lütkepohl (2008) propose an identification scheme that builds upon time heterogeneous covariance estimators. Herwartz and Lütkepohl (2014) show how such heteroskedasticity-based identification approaches can be combined with external information gathered from economic theory. Herwartz and Plödt (2016a) compare the theory based (i.e., using sign restrictions) and data driven (i.e., heteroskedasticity-based) identification schemes by means of a simulation study, and also suggest to combine the two approaches to benefit from complementary information.

More recently developed statistical identification procedures build upon non-normality of the structural shocks on the basis of independent component analysis (Lanne et al., 2017; Gouriéroux et al., 2017; Moneta et al., 2013). Under a non-Gaussian distribution independent (i.e. orthogonal) shocks can be uniquely identified (Comon, 1994).[1] Moneta et al. (2013) have adopted independent component analysis to determine optimal variable orderings in recursive systems of non-Gaussian structural shocks. However, their a-priori focus on triangular schemes appears restrictive in an economic context.

In this study, we compare three alternative independence based identification procedures which allow for more general (non-triangular) transmission schemes: a parametric approach applying ML estimation (Lanne et al., 2017) and two nonparametric procedures based on dependence diagnostics (Herwartz, 2015; Matteson and Tsay, 2013). Similar to Herwartz and Plödt (2016a) we contrast the estimates obtained by sign restrictions with those of independence based identification under a data generating process grounding on an economically motivated dynamic model. We further differentiate the alternative independence based identification schemes and highlight their performance characteristics. In an application to the global crude oil market we find economically reasonable impulse responses based on independence based identification. This outcome supports the results of former studies (e.g., Kilian and Murphy, 2012) where theory based assumptions have been imposed. In general, the data driven identification approaches appear particularly promising if economic assumptions cannot be stated unambiguously.

In the next section, we describe the four alternative identification schemes. Section 3 provides the simulation setting and the corresponding results. Section 4 includes the empirical analysis of the global crude oil market. Section 5 concludes.

2 Identification procedures for structural VAR analysis

Consider a K-dimensional vector autoregressive model of order p

$$\begin{aligned} y_t &= c_t + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \\ &= c_t + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \mathsf{B}\varepsilon_t, \\ \Leftrightarrow \quad A(L) y_t &= c_t + \mathsf{B}\varepsilon_t, \qquad t = 1, \ldots, T, \end{aligned} \tag{1}$$

[1] Note that notions of stochastic dependence and correlation coincide under Gaussianity. Furthermore, the assumption of non-Gaussianity might be reasonable for economic data allowing, e.g., leptokurtic distributions (see, e.g., Chib and Ramamurthy, 2014; Cúrdia et al., 2014, for DSGE models with t-distributed shocks).

with vector valued deterministic terms c_t and $A(L) = I_K - A_1L - \ldots - A_pL^p$, where I_K denotes the $K \times K$ identity matrix. Furthermore, we assume the model to be causal, i.e., $\det(A(z)) \neq 0$ for all $|z| \leq 1$. The stochastic model components are commonly characterized from two perspectives: Firstly, *reduced form residuals* u_t correspond to error terms with zero mean $E(u_t) = 0$ and covariance matrix $\Sigma_u = \mathsf{B}\mathsf{B}'$. Secondly, *structural shocks* $\varepsilon_t = \mathsf{B}^{-1}u_t$ are uncorrelated with $E(\varepsilon_t) = 0$ and $\Sigma_\varepsilon = \mathsf{B}^{-1}\Sigma_u\mathsf{B}^{-1} = I_K$. The factor B of the covariance matrix Σ_u is not unique. However, the matrix B has to be identified properly to allow a meaningful interpretation of the impact of structural shocks. The literature on SVAR models incorporates approaches to solve this identification problem assuming either statistical or economic properties of the structural shocks (for an up-to-date textbook treatment of SVARs see Kilian and Lütkepohl, 2017).

In the following, we first briefly resume the widely applied identification procedure based on sign restrictions. Furthermore, three approaches based on non-normality of structural shocks are considered. The first procedure has been advocated in Lanne et al. (2017) and is based on maximum likelihood (ML) estimation assuming, for instance, t-distributed structural shocks. Relaxing the strict distributional assumptions required for ML estimation, two further identification strategies allow an interpretation as Hodges Lehman (HL, Hodges and Lehmann, 2006) estimation of the structural model.[2] Principles of HL estimation motivate the detection of least dependent structural shocks by the minimization of two alternative nonparametric dependence criteria, namely the so-called distance covariance (dCov) of Székely et al. (2007) and the Cramér-von Mises (CvM) distance of Genest and Rémillard (2004). While the former has already been employed in the context of independent component analysis (Matteson and Tsay, 2013), the latter has been suggested for point estimation of cyclic SVARs by Herwartz (2015). Herwartz and Plödt (2016b) used the CvM criterion for a structural model of the crude oil market.[3]

[2]Conditional on a particular nuisance free test statistic, the HL estimator of a parameter of interest is the specific parameter value obtaining the largest p-value when subjected to testing.

[3]Analyzing the global crude oil market, Herwartz and Plödt (2016b) illustrate that independence criteria give rise to well distinguished supply, general demand, and oil specific demand shocks. Interestingly, these HL estimates closely resemble their counterparts in Kilian and Murphy (2012) that rely on a combination of sign restrictions and further economically motivated inequality patterns.

2.1 Identification based on sign restrictions

Throughout the literature on SVAR models, several variants of identification based on sign restrictions have been applied. These all build upon restricting the structural parameter matrix B to have an economically reasonable sign pattern.[4]

In this study, we consider a stylized version based on least squares estimation of the covariance matrix $\widehat{\Sigma}_u$. After the estimation step, multiplying a rotation matrix Q to the lower triangular Choleski factor D of $\widehat{\Sigma}_u$ generates possible covariance decompositions. Consequently, $\widehat{\Sigma}_u = DD' = DQ(DQ)'$ corresponds to the factorization of the estimated covariance matrix. The matrix Q is a product of Givens rotation matrices defined through the corresponding $(K(K-1)/2) \times 1$ dimensional vector of rotation angles $\theta = (\theta_1, \ldots, \theta_{K(K-1)/2})$ to cover the entire space of covariance decompositions (for more technical details, see, for instance, Canova and Nicolo, 2002).

Identification by means of sign restrictions consists of generating a large set of the $Q(\theta)$ matrices by drawing the rotation angles θ_i, $i = 1, \ldots, K(K-1)/2$, uniformly from the interval $[0, \pi]$. Accordingly, this draw is successful if the associated decomposition $\mathsf{B}(\theta) = D\,Q(\theta)$ fulfills the a-priori specified sign pattern. The sampling proceeds until a prespecified number of successful draws (10,000, say) has been obtained. Often, the identified matrix, henceforth denoted $\widehat{\mathsf{B}}_{\mathrm{SR}}$, corresponds to the median of this subset of all matrix candidates showing the preselected sign pattern (see Fry and Pagan, 2007, for a critique of this convention).

2.2 Independence based identification

We describe three identification procedures which build upon the assumption of non-Gaussian structural shocks. More precisely, following these approaches the vector of structural shocks ε_t is allowed to contain at most one normally distributed component $\varepsilon_{t,k}$. For non-Gaussian vector valued shocks, the identification problem introduced above coincides with the aim of

[4]As typical elements of B the parameters b_{ij} quantify direction and magnitude of the contemporaneous effect of a (positive) structural unit shock ε_{jt} on variables y_{it}. For both characteristics - direction and (relative) magnitude - economic theory might offer plausible arguments. Against this background, distinguished procedures can be applied to derive the corresponding estimate of the matrix B (see, for instance, Uhlig, 2005; Fry and Pagan, 2007). The stylized sign restrictions considered in this study formalize throughout directional effects.

independent component analysis (ICA). Following the fundamental result of Comon (1994), ICA identifies the so-called mixing matrix B which is unique up to column signs and ordering. The resulting vector of structural shocks $\varepsilon_t = \mathsf{B}^{-1}u_t$ consists of independently distributed components. The literature on ICA comprises several approaches and algorithms to determine B. The basic ICA procedures, for instance described in Hyvärinen et al. (2010), build upon an assumption of acyclicity in the causal scheme of the variables such that the resulting matrix B is lower triangular. These methods are applied to SVAR models within the VAR-LiNGAM (linear, non-Gaussian, acyclic model) algorithm of Moneta et al. (2013). The procedures considered below are more general in the sense that they allow for cyclic causality and hence, appear more appropriate in macroeconomic applications.[5]

After estimation of the model, structural shocks are retrieved from estimated reduced form disturbances $\hat{u}_t$ as $\hat{\varepsilon}_t = \mathsf{B}^{-1}\hat{u}_t$. Next, we briefly sketch ML and HL estimation (based on the Cramér-von Mises statistic and the distance covariance) of structural shocks $\hat{\varepsilon}_t$.

2.2.1 Identification based on ML estimation

Lanne et al. (2017) suggest to determine B by means of maximization of the joint density of independent non-normally distributed variables. Let f_k denote the densities of components $\varepsilon_{t,k}$, $k = 1, \ldots, K$. Respective distributional parameters are collected in λ_k.[6] Furthermore, to have a unit diagonal the matrix of structural parameters $\mathsf{B}(\beta)$ is columnwise normalized by the corresponding standard deviation σ_k. The vector β collects the vectorized off-diagonal elements of the standardized matrix $\mathsf{B}(\beta)$.

With these conventions ML estimation of B proceeds in two steps. In the first step, least squares estimates $\hat{u}_t$ are estimated from the VAR model.[7] Based on the estimated $\hat{u}_t$, the

[5] For instance, results of Herwartz and Lütkepohl (2014) and Herwartz and Plödt (2016a) are at odds with a recursive causation scheme in three and four dimensional SVAR models of US monetary policy.

[6] Note that the component densities f_k each depend on (possibly distinct) parameter values λ_k which can, for instance, correspond to the family of t-distributions with λ_k degrees of freedom (Lanne et al., 2017).

[7] We apply the two-step ML estimation procedure rather than simultaneous estimation of the residuals and structural parameters which might be rather demanding even for medium dimensions K and time series of moderate length (Lanne et al., 2017).

log-likelihood

$$L_T(\beta, \sigma, \lambda) = T^{-1} \sum_{t=1}^{T} l_t(\beta, \sigma, \lambda), \tag{2}$$

where

$$l_t(\beta, \sigma, \lambda) = \sum_{k=1}^{K} \log f_k(\sigma_k^{-1} \iota_k' \mathsf{B}(\beta)^{-1} \hat{u}_t; \lambda_k) - \log \det(\mathsf{B}(\beta)) - \sum_{k=1}^{K} \log \sigma_k \tag{3}$$

is maximized with respect to the parameter vector $(\beta', \sigma, \lambda)'$ comprising standard deviations $\sigma = (\sigma_1, \ldots, \sigma_K)$, the structural parameter vector β and the component specific distribution parameters $\lambda = (\lambda_1', \ldots, \lambda_K')'$. After multiplication with the transposed k-th unit vector, denoted ι_k', the structural shocks are $\hat{\varepsilon}_{t,k} = \sigma_k^{-1} \iota_k' \mathsf{B}(\beta)^{-1} \hat{u}_t$. With ML estimates $\tilde{\sigma}$ and $\tilde{\beta}$ the matrix B is estimated as $\widehat{\mathsf{B}}_{\mathrm{ML}} = \mathsf{B}(\tilde{\beta})\mathrm{diag}(\tilde{\sigma})$. It is noteworthy that the identification procedure in Gouriéroux et al. (2017) can be seen as an extension of this parametric approach.

2.2.2 Identification based on the distance covariance (dCov)

ML estimation (Lanne et al., 2017) proceeds under the assumption of a well and fully specified distributional framework. Theoretical and asymptotic properties of quasi ML estimators, i.e., estimators maximizing a misspecified Gaussian log-likelihood, have attracted interest in several branches of econometric literature. The mere existence of this literature highlights that a-priori distributional assumptions might be difficult to justify in practice. Furthermore, it is worth to point out that maximizing a non-Gaussian likelihood under misspecification in the present context is not well understood and bears severe risks of first and second order estimation biases (see, e.g., Newey and Steigerwald, 1997, for the case of non-Gaussian quasi ML estimation in GARCH models). Avoiding any restrictive assumption on the distribution of ε_t, nonparametric dependence measures can be applied alternatively for identification. Implementing ICA, Matteson and Tsay (2013) apply a nonparametric dependence measure, namely the so-called distance covariance of Székely et al. (2007). The ICA algorithm provides a matrix estimate $\widehat{\mathsf{B}}$ such that the corresponding structural shocks $\hat{\varepsilon}_t = \widehat{\mathsf{B}}^{-1} \hat{u}_t$ minimize the distance covariance, i.e., are least dependent according to this statistic. Similar to the procedure based on sign restrictions, the set of possible structural matrices $\mathsf{B}(\theta) = D\, Q(\theta)$ is defined in terms of the Choleski factor D and the vector of rotation angles θ of the Givens matrices $Q(\theta)$. Accordingly, the distance covariance $\mathcal{U}_T(\hat{\varepsilon}_t(\theta))$ can be calculated from $\hat{\varepsilon}_t(\theta) = \mathsf{B}(\theta)^{-1} \hat{u}_t$. In the sense of HL estimation, the distance covariance $\mathcal{U}_T$ is minimized by $\hat{\theta} =$

$\operatorname{argmin}_\theta \mathcal{U}_T(\hat{\varepsilon}_t(\theta))$ which consequently determines the estimated matrix $\widehat{\mathsf{B}}_{\text{dCov}} = \mathsf{B}(\hat{\theta})$.

For details on the exact minimization procedure and the empirical definition of the dependence measure we refer to the Appendix and Matteson and Tsay (2013). In this study, we apply the function *steadyICA* from the `R` package `steadyICA` (Risk et al., 2015) to determine $Q(\hat{\theta})$.

2.2.3 Identification based on the Cramér-von Mises statistic (CvM)

There are diverse nonparametric criteria to measure the degree of dependence between random variables, one of which, namely the distance covariance, was described in the last subsection. Besides this, Genest and Rémillard (2004) introduce a nonparametric criterion to quantify the dependence within a set of random variables based on the CvM statistic $\mathcal{C}$ (the exact definition of $\mathcal{C}$ is given in the Appendix). This statistic quantifies the distance between the empirical copula and the independence copula of the components of the random sample $\hat{\varepsilon}_t$ for $t = 1, \ldots, T$. As an alternative to using the distance covariance, least dependent shocks $\hat{\varepsilon}_t$ are then characterized by a minimal distance in the CvM sense.

Similar to the procedure of Matteson and Tsay (2013), the statistic $\mathcal{C}(\hat{\varepsilon}_t(\theta))$ is minimized with respect to a vector of rotation angles θ for alternative starting values θ_0.[8] For $\hat{\theta} = \operatorname{argmin}_\theta \mathcal{C}(\hat{\varepsilon}_t(\theta))$, the estimated matrix $\widehat{\mathsf{B}}_{\text{CvM}} = \mathsf{B}(\hat{\theta})$ corresponds to the least dependent structural shocks $\hat{\varepsilon}_t(\hat{\theta})$. We use the implementation of $\mathcal{C}$ in the `R` package `copula` (Hofert et al., 2015). An alternative dependence criterion is, for instance, applied in Capasso and Moneta (2016).[9]

[8] All considered identification schemes base on optimization techniques. While we apply an implemented R function for minimization of the distance covariance, we find the minimum of the CvM criterion based on several starting values. Furthermore, the ML method incorporates a relatively large parameter vector $(\beta', \sigma, \lambda)$. It is noteworthy that the identification relying on independent structural shocks are of comparable computational complexity.

[9] Apart from computational merits, targeting structural shocks showing weakest dependence in terms of the CvM diagnostic holds the advantage that $\mathcal{B}$ is consistent against any form of dependence (Genest and Rémillard, 2004). Additionally, the two dependence criteria applied here outperform alternative dependence measures in terms of power against a wide range of dependence structures (Herwartz and Maxand, 2017). Matteson and Tsay (2013) show by means of a simulation study that the ICA algorithm employed in Capasso and Moneta (2016) (*FastICA*) invokes larger mean errors compared with the algorithm implemented in *steadyICA*.

3 Simulation study

Applying the most suitable identification procedure might be crucial for the estimation of structural matrices. The following simulation study sheds light on the performance of the four identification schemes described in Section 2. Specifically, we compare the estimated matrices $\widehat{\mathsf{B}}_{\bullet}$, $\bullet \in \{\text{SR, ML, dCov, CvM}\}$, under alternative distributional settings of the structural shocks. Subsection 3.1 describes how the structural shocks are generated followed by details on the performance evaluation in 3.2. In Subsection 3.3 we discuss the simulation results.

3.1 Data generation

Reduced form residuals u_t are generated by means of an economically reasonable simulation framework as described in Herwartz and Plödt (2016a). Simulated data generating processes (DGPs) resemble a stylized three-equation dynamic stochastic general equilibrium (DSGE) model comprising the output gap (x_t), inflation (π_t) and nominal interest rates r_t (Gertler et al., 1999; Carlstrom et al., 2009; Castelnuovo, 2013, 2012, 2016). The model consists of a New Keynesian IS equation, a hybrid New Keynesian Phillips curve, and a Taylor rule with interest rate smoothing. First order autoregressive innovations characterize demand, supply and monetary policy shocks. The resulting three dimensional SVAR reads as

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \mathsf{B}\varepsilon_t, \quad t = 1, \dots, T, \tag{4}$$

where $y_t = (x_t, \pi_t, r_t)'$. Based on typical calibrations of the underlying DSGE model, the associated autoregressive matrices A_1, A_2 and the structural parameter matrix B are[10]

$$A_1 = \begin{pmatrix} 1.24 & -0.09 & -0.16 \\ 0.13 & 0.94 & -0.06 \\ 0.24 & 0.30 & 1.03 \end{pmatrix}, A_2 = \begin{pmatrix} -0.37 & 0.05 & 0.08 \\ -0.07 & -0.22 & 0.03 \\ -0.12 & -0.15 & -0.27 \end{pmatrix}, \mathsf{B} = \begin{pmatrix} 2.32 & -0.48 & -0.41 \\ 0.72 & 2.32 & -0.22 \\ 0.98 & 1.57 & 0.76 \end{pmatrix}. \tag{5}$$

[10] For a detailed description of the underlying DSGE model and a detailed motivation of the parametrization see Herwartz and Plödt (2016a) and the Appendix. The derivation of the log linearized SVAR from the underlying DSGE model relies on a first order Taylor series expansion. Under deviations from Gaussian innovations the numerical solution provided in (5) is not robust with regard to higher order moments of the true shocks, i.e., with regard to the solution of higher order Taylor series expansions. Since our interest is not in the most accurate dynamic description of economic optimization solutions and to avoid distribution specific DGPs, we abstract from this point to take advantage of scenario independent 'true' parameter values in simulated DGPs.

The matrix B implies a unique pattern of instantaneous effects of the shocks on the variables.[11] Demand shocks invoke an increase in all three variables on impact while a positive supply shock raises the levels of prices and interest rates but implies a negative response of output. A positive interest rate shock mutes inflation and dampens economic activity. It is noteworthy that the monetary policy shock is already identified by means of its presumed counter directional impact on policy rates and prices. Therefore, the effect of monetary policy on output is often left unspecified in the literature on sign restrictions (Uhlig, 2005).

Furthermore, we assume homoskedastic structural shocks $\varepsilon_t \sim iid(0, I)$ in (4). More precisely, the elements of the trivariate vectors, $\varepsilon_t = (\varepsilon_{x,t}, \varepsilon_{\pi,t}, \varepsilon_{r,t})$, are drawn as independent, identically distributed univariate variables following standardized Student-t distributions with alternative degrees of freedom, $\nu = 5, 10, 20$, and a centered and standardized χ^2-distribution with $\nu = 5$ degrees of freedom. With these vector valued structural shocks ε_t and the matrices A_1, A_2 and B given in (5), we generate samples $\{y_t\}_{t=-100}^{T}$ of size $T = 100, 250, 500$ according to (4).[12] From the generated processes, we estimate least squares residuals $\hat{u}_t$ assuming that the true autoregressive order ($p = 2$) is known to the analyst. Conditional on sample information $\{\hat{u}_t\}_{t=1}^{T}$ we calculate the structural matrix B by means of the alternative procedures described in Section 2 obtaining $\widehat{\mathsf{B}}_\bullet$, $\bullet \in \{$SR, ML, dCov, CvM$\}$.

3.2 Performance evaluation

In the first place, we evaluate the performance of alternative identification schemes by means of the mean squared error (MSE) of the estimated matrices with respect to the true transmission matrix B given in (5). More precisely, we quantify the distance between the estimates $\widehat{\mathsf{B}}_{\bullet,l}$ and B in terms of the Frobenius norm $||\cdot||_F$ of their difference. Let l, $l = 1, 2, \ldots, L$, $L =$ 1000 denote an indexation of single Monte Carlo experiments. Then, for matrices $\widehat{\mathsf{B}}_{\bullet,l}$ and

[11] Herwartz and Plödt (2016a) show that the response pattern implied by B is robust under a broad set of parameter calibrations of the underlying DSGE model.

[12] After the generation step we drop the first 100 observations to immunize simulation results against the effects of initial conditions.

identification scheme $\bullet \in \{\text{SR, ML, dCov, CvM}\}$ the MSE of $\widehat{\mathsf{B}}_\bullet$ is

$$\widehat{MSE}_\bullet = \frac{1}{L}\sum_{l=1}^{L} \inf_{P\in\mathcal{P}} ||\mathsf{B} - \widehat{\mathsf{B}}_{\bullet,l}P||_F$$
$$= \frac{1}{L}\sum_{l=1}^{L} \frac{1}{K} \inf_{P\in\mathcal{P}} \sqrt{\sum_{i=1}^{K}\sum_{j=1}^{K}\Big(\mathsf{B}_{ij} - (\widehat{\mathsf{B}}_{\bullet,l}P)_{ij}\Big)^2}. \tag{6}$$

Noticing that identification outcomes are 'unique' up to column signs and column ordering, the infimum in (6) is taken over all matrices $P \in \mathcal{P}$ of the subset of signed column permutation matrices of $\widehat{\mathsf{B}}_{\bullet,l}$ within the set of the nonsingular $K \times K$ matrices. For more details on the determination of the infimum we refer to the R documentation of the function $frobICA$ (Risk et al., 2015). Note that the formulation in (6) neglects the non-uniqueness of the estimated matrices with respect to signed permutations by choosing the matrix P which fits best to the true matrix of structural parameters B. Thus, we assume that a proper matrix P can be identified, for instance, based on economic theory.

Secondly, we compare the identification procedures based on the independence of non-Gaussian structural shocks by reporting selected statistics summarizing the sign patterns of the estimated matrices $\widehat{\mathsf{B}}_\bullet$, $\bullet \in \{\text{ML, dCov, CvM}\}$. To be specific, for the case of the considered three dimensional model we display all possible identified sign patterns in Table 1 and indicate them with roman numbers I to VI.[13] For each Monte Carlo experiment we determine the sign pattern of $\widehat{\mathsf{B}}_{\bullet,l}$, $\bullet \in \{\text{ML, dCov, CvM}\}$, and report the resulting frequencies of sign patterns I,...,VI. In addition, a structural estimate might lack overall identification, i.e., the estimated sign pattern lacks uniqueness. We indicate unidentified sign patterns with '0' and also provide strategy specific frequencies of overall unidentified estimates. Since the identified matrices $\widehat{\mathsf{B}}_{\text{SR}}$ hold the true sign pattern IV by assumption, we do not include identification by means of (correct) sign restrictions in this direction of performance assessment.

[13]The displayed matrices are representative for all sign patterns that can be determined by signed permutations of the columns. In contrast, we consider a matrix as not identified if the sign pattern of at least two columns of the matrix is identical or can be obtained by means of multiplication with –1. Note that we restrict the diagonal of B to have a positive sign to focus on the effects of positive structural shocks.

I			II			III			**IV**			V			VI		
+	−	−	+	−	−	+	+	−	+	−	−	+	−	−	+	−	+
−	+	−	+	+	−	+	+	−	+	+	−	+	+	+	+	+	−
+	+	+	+	−	+	+	−	+	+	+	+	+	−	+	+	−	+

Table 1: Identified sign patterns in case $K = 3$ (unidentified sign patterns are indicated with '0'); sign pattern IV corresponds to the true sign pattern of matrix B defined in (5).

3.3 Simulation results

We first highlight performance differences of the identification procedures in terms of the MSE defined in (6). Additionally, the frequency of sign patterns of the estimated matrices $\widehat{\mathsf{B}}_\bullet$, $\bullet \in \{\text{ML, dCov, CvM}\}$ reflect the performance in comparison with economically reasonable directions of the impact of shocks (a-priori implemented within the sign restrictions).

Mean squared errors

Table 2 documents the mean squared errors $\widehat{MSE}_\bullet$, $\bullet \in \{\text{SR, ML, dCov, CvM}\}$, with respect to the distribution chosen for data generation and alternative sample sizes. In the following, we first discuss the results for identification based on sign restrictions and subsequently, the MSE estimates for independence based identification.

Throughout, the estimated mean squared errors $\widehat{MSE}_{\text{SR}}$ indicate a bias in estimating B by means of sign restrictions. In particular, these results reveal that the bias neither depends on the underlying distribution nor vanishes for increasing sample sizes, i.e., the estimation remains biased even asymptotically. Noticing that the application of sign restrictions can be seen as a censored sampling from the set of possible covariance decomposition matrices B, an estimation bias naturally arises in this context.[14]

[14] Reducing the set-identification to the median matrix (or distinct indices, see Fry and Pagan, 2011) allows to quantify the occurring bias for $\widehat{\mathsf{B}}_{\text{SR}}$. Note that we consider a relatively simple implementation of the sign restriction approach where the model is fully restricted. The bias can be reduced by incorporating further characteristics of the data prior to estimation, for instance, in the framework of a more agnostic model (Arias et al., 2014).

	T	SR	ML	dCov	CvM
$t(5)$	100	0.3098	0.2373	0.2464	0.3397
	250	0.3068	0.1441	0.1768	0.3045
	500	0.3062	0.0965	0.1260	0.2512
$t(10)$	100	0.3087	0.3138	0.2861	0.3667
	250	0.3062	0.3133	0.2689	0.3660
	500	0.3052	0.1950	0.2249	0.3467
$t(20)$	100	0.3086	0.3418	0.3028	0.3712
	250	0.3062	0.3133	0.2898	0.3659
	500	0.3056	0.2820	0.2711	0.3647
$\chi^2(5)$	100	0.3096	0.2748	0.1502	0.2211
	250	0.3063	0.1965	0.0846	0.1244
	500	0.3061	0.1325	0.0563	0.0826

Table 2: $\widehat{MSE}_{\bullet}$ for $\bullet \in \{\text{SR, ML, dCov, CvM}\}$ calculated as in (6) from $L = 1000$ Monte Carlo experiments.

MSE estimates $\widehat{MSE}_{\bullet}$, $\bullet \in \{\text{ML, dCov, CvM}\}$, vary with both the sample size and the underlying distribution. The performance of all identification schemes that build upon the assumption of independent non-Gaussian shocks (ML, dCov and CvM) improves with increasing sample sizes, i.e., the identified matrices $\widehat{\mathsf{B}}_{\bullet}$ approach the true parameters. Apparently the independence based identification benefits from consistency of the independence diagnostics and ML estimation under correct likelihood specification.[15] Evaluating the performance of the identification schemes for t-distributed structural shocks, we diagnose performance differentials for distinct degrees of freedom: For lower degrees of freedom MSE estimates are smaller. In addition to less precise estimation for larger degrees of freedom, the corresponding MSE estimates decrease only mildly with increasing sample sizes in the case of $\nu = 20$. These results reflect that the matrix B can only be uniquely determined for non-Gaussian shocks.

[15]Under misspecification of the log likelihood one cannot exclude (asymptotic) estimation biases such that convergence to the true matrix of structural parameters might not apply.

Especially in small samples it appears difficult to distinguish Gaussian samples from shocks exhibiting a standardized Student-t distribution with 20 degrees of freedom.

ML estimates $\widehat{\mathsf{B}}_{\mathrm{ML}}$ depend on a prespecified distribution family (the t-distribution in our case). Unsurprisingly, the ML estimator performs best for all sample sizes for processes drawn from Student-t distributed shocks with $\nu = 5$. With larger degrees of freedom, however, the HL estimator based on the distance covariance, $\widehat{\mathsf{B}}_{\mathrm{dCov}}$, outperforms the rival identification procedures in small samples ($T = 100$). In larger samples ($T = 500$), the ML method shows the smallest MSE. Throughout, the HL estimators $\widehat{\mathsf{B}}_{\mathrm{CvM}}$ and $\widehat{\mathsf{B}}_{\mathrm{dCov}}$ perform similarly with the former obtaining larger MSE estimates.

Under χ^2-distributed structural shocks MSE estimates of all independence based identification approaches shrink for increasing sample sizes. In this setting the MSEs of the nonparametrically identified estimates $\widehat{\mathsf{B}}_{\mathrm{CvM}}$ or $\widehat{\mathsf{B}}_{\mathrm{dCov}}$ are smaller than the MSE of $\widehat{\mathsf{B}}_{\mathrm{ML}}$ maximizing the misspecified Student-t likelihood. Furthermore, these identification procedures ($\mathsf{B}_{\mathrm{dCov}}$, $\mathsf{B}_{\mathrm{CvM}}$) substantially improve compared with the results for data drawn from t-distributed shocks.

Sign patterns

In Tables 3 and 4 we document the frequency of estimated matrices $\widehat{\mathsf{B}}_{\bullet}$, $\bullet \in \{\mathrm{ML}, \mathrm{dCov}, \mathrm{CvM}\}$, that accord with the respective sign patterns 0, I,..., VI (see Table 1) under $t(5)$-distributed and $\chi^2(5)$-distributed structural shocks, respectively. The results are generally in line with those discussed for the estimated MSEs. With increasing sample size the frequency of matrices $\widehat{\mathsf{B}}_{\bullet}$ featuring the correct sign pattern increases.

In presence of Student-t distributed shocks with $\nu = 5$ maximizing the (correctly specified) likelihood obtains the highest frequency of correct sign patterns. In 86.5% of all experiments with sample size $T = 500$ the sign pattern of $\widehat{\mathsf{B}}_{\mathrm{ML}}$ accords with the true sign pattern IV. For a sample of this size, identification based on dCov (CvM) obtains matrices with correct sign patterns in 72.8% (42.0%) of all experiments. The sign pattern with second highest frequency (or highest in case of CvM identification and samples of size $T = 250, 500$) is sign pattern 0, i.e., the group of overall unidentified structural estimates. Among the identified matrices

$t(5)$	T	0	I	II	III	**IV**	V	VI
ML	100	0.358	0.010	0.039	0.000	**0.462**	0.128	0.003
	250	0.239	0.001	0.008	0.000	**0.712**	0.040	0.000
	500	0.120	0.000	0.000	0.000	**0.865**	0.015	0.000
dCov	100	0.486	0.008	0.044	0.000	**0.293**	0.159	0.010
	250	0.322	0.003	0.031	0.000	**0.548**	0.095	0.001
	500	0.217	0.000	0.009	0.000	**0.728**	0.046	0.000
CvM	100	0.551	0.006	0.035	0.000	**0.217**	0.185	0.006
	250	0.504	0.004	0.046	0.000	**0.283**	0.159	0.004
	500	0.413	0.002	0.030	0.000	**0.420**	0.133	0.002

Table 3: Frequency of matrices $\widehat{\mathsf{B}}_\bullet$, $\bullet \in \{\text{ML, dCov, CvM}\}$ holding the corresponding sign pattern based on $t(5)$-distributed structural shocks and $L = 1000$ (identified sign patterns are indicated with I to VI and overall unidentified sign patterns with 0).

almost all estimated outcomes are concentrated in sign patterns IV and V.[16]

In contrast, for $\chi^2(5)$-distributed structural shocks identification based on the misspecified likelihood provides less correctly identified sign patterns. For samples of size $T = 500$ the correct sign pattern is identified in 70.9% of all experiments. In this setting, the performance of identification based on CvM and dCov substantially improves in comparison with the case of data drawn from t-distributed structural shocks. For $T = 500$ the frequencies to detect the correct sign pattern are 90.1% and 97.8% for the CvM and dCov based identification, respectively.

In summary, the MSE and the sign patterns of structural estimates $\widehat{\mathsf{B}}_\bullet$ highlight performance differentials among the identification schemes $\bullet \in \{\text{SR, ML, dCov, CvM}\}$. The consideration of scenarios with alternative distributions and sample sizes allows to quantify the bias which occurs for the sign restriction approach throughout. In contrast to sign restrictions, the performance of independence based identification improves for increasing sample sizes

[16] After column permutation and sign change the sign pattern in V also comprises economically reasonable effects, with a monetary policy shock fostering output.

$\chi^2(5)$	T	0	I	II	III	**IV**	V	VI
ML	100	0.411	0.014	0.045	0.000	**0.372**	0.153	0.005
	250	0.315	0.004	0.024	0.000	**0.562**	0.093	0.002
	500	0.243	0.000	0.005	0.000	**0.709**	0.043	0.000
dCov	100	0.258	0.001	0.005	0.000	**0.671**	0.065	0.000
	250	0.101	0.000	0.000	0.000	**0.892**	0.007	0.000
	500	0.022	0.000	0.000	0.000	**0.978**	0.000	0.000
CvM	100	0.389	0.006	0.018	0.000	**0.443**	0.139	0.005
	250	0.183	0.002	0.001	0.000	**0.753**	0.061	0.000
	500	0.083	0.000	0.000	0.000	**0.901**	0.016	0.000

Table 4: Frequency of matrices $\widehat{\mathsf{B}}_\bullet$, $\bullet \in \{\text{ML, dCov, CvM}\}$ holding the corresponding sign pattern based on $\chi^2(5)$-distributed structural shocks and $L = 1000$ (identified sign patterns are indicated with I to VI and overall unidentified sign patterns with 0).

where the ML method can be distinguished from the nonparametric identification schemes. Our results allow a general conclusion: If the distribution family for the maximum likelihood method is specified correctly, ML estimation performs best (at least in large samples). Otherwise nonparametric identification schemes show a superior and more reliable performance. In particular, in terms of MSE and the frequency of correct sign patterns identification based on the distance covariance outperforms the HL estimator that is based on the Cramér-von Mises statistic.

4 A model of the global crude oil market

When analyzing the effects of oil price fluctuations on macroeconomic aggregates several types of oil shocks are commonly distinguished in the literature. For instance, Kilian and Murphy (2012) focus on disentangling oil supply and demand shocks on the real price of oil. They apply theory based sign restrictions combined with elasticity bounds to identify the oil shocks (see, for instance, Peersman and Van Robays, 2012, for a similar approach). Furthermore, Lütkepohl and Netšunajev (2014) and Herwartz and Plödt (2016b) refrain from using theory based restrictions but rather apply statistical identification based on a change in volatility and the independence of structural shocks, respectively.

To decide on the relative importance of oil supply and demand shocks on the variables of the oil market, we examine a common VAR model of the global crude oil market as considered in Kilian (2009). More precisely, we employ the $K = 3$ dimensional VAR model formulation of Herwartz and Plödt (2016b). Applying identification based on CvM statistics they provide impulse responses comparable to those in Kilian and Murphy (2012).

4.1 The SVAR model

For the model formulated as in (1) the vector $y_t = (\Delta q_t, x_t, p_t)'$ includes the change in global crude oil production, Δq_t, a measure of real economic activity, x_t, and the real price of oil, p_t. The vector ε_t comprises a set of structural shocks, typically labeled as oil supply shock (ε_s), aggregate demand shock (ε_{ad}), and oil-specific demand shock (ε_{osd}), which are uncorrelated across equations and over time with mean zero and unit covariance matrix. Economic arguments imply a clear sign pattern with regard to the (on impact) effects of the shocks on the variables in the system (see, for instance, Peersman and Van Robays, 2012). Table 5 summarizes the theoretical sign pattern, which also forms the basis for an imposition of stylized sign restrictions. In accordance with this sign pattern, we normalize the three shocks in the subsequent analysis such that they raise the real price of oil on impact.

	Shock		
Variable	$\varepsilon_s \rightarrow$	$\varepsilon_{ad} \rightarrow$	$\varepsilon_{osd} \rightarrow$
q	$-$	$+$	$+$
x	$-$	$+$	$-$
p	$+$	$+$	$+$

Table 5: Theoretical sign pattern of shocks in the global crude oil market.

The estimation of the SVAR model is based on monthly data for the period 1973:M1 to 2014:M12.[17] As in Herwartz and Plödt (2016b) we obtain estimates for the matrices

[17] Herwartz and Plödt (2016b) provide a detailed description of the data sources and transformations. They

$A_1, \ldots, A_p$ and the residuals u_t in (1) by OLS estimation of the reduced form VAR model including $p = 24$ lags.

Within this framework, we consider the impact of the structural shocks identified by means of the four identification schemes described in Section 2. To examine the instantaneous and dynamic feedback relations, standard errors and confidence intervals obtain from evaluating so-called fixed design wild bootstrap samples (Gonçalves and Kilian, 2004)

$$y_t^* = \hat{c}_t + \hat{A}_1 y_{t-1} + \hat{A}_2 y_{t-2} + \cdots + \hat{A}_p y_{t-p} + u_t^*, \quad t = 1, \ldots, T. \tag{7}$$

In (7) $\hat{A}_j$, $j = 1, \ldots, p$, and c_t are OLS parameter estimates retrieved from the data. For bootstrap errors $u_t^* = w_t \hat{u}_t$, the scalar random variable w_t exhibits a Rademacher distribution which is independent of the data, i.e., with probability 0.5 it is either unity or minus unity. For error terms $\hat{u}_t^*$ estimated from (7) we determine the bootstrap covariance decomposition by $\widehat{\mathsf{B}}_\bullet^{**} = \widehat{\Sigma}_u^{1/2} \widehat{\Sigma}_{\hat{u}^*}^{-1/2} \widehat{\mathsf{B}}_\bullet^*$. Here, $\widehat{\mathsf{B}}_\bullet^*$ corresponds to the decomposition of $\widehat{\Sigma}_{\hat{u}^*}$ based on identification procedure $\bullet \in \{$SR, ML, dCov, CvM$\}$. The matrices $\widehat{\Sigma}_u^{1/2}$ and $\widehat{\Sigma}_{\hat{u}^*}^{1/2}$ are symmetric eigenvalue decompositions of $\widehat{\Sigma}_u$ and $\widehat{\Sigma}_{\hat{u}^*}$, respectively. Thus, $\widehat{\mathsf{B}}_\bullet^{**}$ decomposes the covariance matrix $\widehat{\Sigma}_u$ allowing to compare the bootstrap decomposition directly with $\widehat{\mathsf{B}}_\bullet$. Subsequently, we determine the order and sign of the columns of $\widehat{\mathsf{B}}_\bullet^{**}$ such that the Frobenius distance (defined in (6)) to $\widehat{\mathsf{B}}_\bullet$ is minimal.

4.2 Empirical results

The independence based identification schemes build on a non-normality assumption of the structural shocks. In support of this assumption, the Jarque-Bera statistics provided in Herwartz and Plödt (2016b) indicate strong evidence against the null hypothesis of normality of the VAR residuals (p-value < 0.001). Moreover, the ML method is based on a t-distributed likelihood. The ML estimates for the degrees of freedom of the t-distribution of the first, second and third structural shock correspond to $df = 3.28, 5.99$ and 5.15, respectively.

In the following, we consider the estimated matrices of contemporaneous effects with corresponding standard errors and the dynamic impact of structural shocks in terms of impulse response functions. Based on the alternative identification schemes the respective estimates of the structural matrix read as

also describe the explicit choice of the sign pattern in more detail.

$$\widehat{\mathsf{B}}_{\mathrm{SR}} = \begin{pmatrix} \underset{[0.348]}{-0.880} & \underset{[0.250]}{0.245} & \underset{[0.364]}{1.000} \\ \underset{[0.211]}{-0.234} & \underset{[0.217]}{0.830} & \underset{[0.278]}{-0.507} \\ \underset{[1.676]}{5.258} & \underset{[1.688]}{4.565} & \underset{[1.578]}{1.435} \end{pmatrix}, \quad \widehat{\mathsf{B}}_{\mathrm{ML}} = \begin{pmatrix} \underset{[0.016]}{-1.433} & \underset{[0.008]}{0.049} & \underset{[0.016]}{-0.037} \\ \underset{[0.009]}{0.025} & \underset{[0.010]}{0.988} & \underset{[0.014]}{-0.125} \\ \underset{[0.017]}{0.313} & \underset{[0.019]}{1.791} & \underset{[0.065]}{6.972} \end{pmatrix},$$

$$\widehat{\mathsf{B}}_{\mathrm{dCov}} = \begin{pmatrix} \underset{[0.008]}{-1.423} & \underset{[0.078]}{0.028} & \underset{[0.117]}{0.015} \\ \underset{[0.057]}{-0.017} & \underset{[0.021]}{1.058} & \underset{[0.160]}{0.078} \\ \underset{[0.625]}{0.810} & \underset{[1.113]}{0.394} & \underset{[0.264]}{7.498} \end{pmatrix}, \quad \widehat{\mathsf{B}}_{\mathrm{CvM}} = \begin{pmatrix} \underset{[0.098]}{-1.321} & \underset{[0.181]}{0.340} & \underset{[0.239]}{0.407} \\ \underset{[0.099]}{-0.109} & \underset{[0.212]}{0.691} & \underset{[0.197]}{-0.798} \\ \underset{[1.333]}{3.329} & \underset{[1.273]}{5.873} & \underset{[1.557]}{3.386} \end{pmatrix}, \qquad (8)$$

where standard errors are given in brackets (calculated from all successful draws for $\widehat{\mathsf{B}}_{\mathrm{SR}}$ and bootstrap samples for $\widehat{\mathsf{B}}_{\bullet}, \bullet \in \{\mathrm{ML, dCov, CvM}\}$).

Although the ordering and signs of the columns of $\widehat{\mathsf{B}}_{\bullet}, \bullet \in \{\mathrm{SR, ML, dCov, CvM}\}$ can not be uniquely determined in statistical terms, economic justifications might lead to an adequate labeling of the resulting structural shocks. Following Kilian and Murphy (2012) we aim at ordering effects of a positive oil supply shock first, of an aggregate demand shock second and an oil-specific demand shock in the third place. For the shock labelling we combine two distinct strategies building on common economic intuitions.[18] More precisely, we rely on the sign pattern implied by the theoretical considerations regarding different types of shocks in the global crude oil market (cf. Table 5) and arguments in Kilian and Murphy (2012). While the estimated matrix $\widehat{\mathsf{B}}_{\mathrm{SR}}$ fulfills the sign pattern in Table 5 by construction, we can reorder (and multiply by -1) the columns of $\widehat{\mathsf{B}}_{\mathrm{CvM}}$ so that this estimate also coincides with the theoretical sign pattern (see (8)). In contrast, on impact effects in $\widehat{\mathsf{B}}_{\mathrm{ML}}$ and $\widehat{\mathsf{B}}_{\mathrm{dCov}}$ only partly conform with the theoretical pattern. Entry $\widehat{\mathsf{b}}_{23,\mathrm{dCov}}$ and entries $\widehat{\mathsf{b}}_{13,\mathrm{ML}}$ and $\widehat{\mathsf{b}}_{21,\mathrm{ML}}$ depart from the economically suspected sign patterns. In particular, the last two columns of $\widehat{\mathsf{B}}_{\mathrm{dCov}}$ lack identification by their sign pattern and can only be identified by the size of their entries. To distinguish the last two columns of $\widehat{\mathsf{B}}_{\mathrm{ML}}$ and $\widehat{\mathsf{B}}_{\mathrm{dCov}}$ we follow Kilian and Murphy

[18] In contrast, Lanne et al. (2017) propose an algorithm to determine completely identified matrices holding a unique (diagonal elements equal to one and unambiguous ordering of the remaining entries) but not necessarily economically reasonable ordering.

(2012) and relate the stronger instantaneous response of the real price of oil to an oil-specific demand shock rather than an aggregate demand shock.

The bootstrap standard errors in (8) indicate that the alternative identification approaches provide structural matrix estimates $\widehat{\mathsf{B}}_\bullet$ with distinct estimation uncertainty. The standard errors of $\widehat{\mathsf{B}}_{\mathrm{ML}}$ appear to be smallest (between 0.08 and 0.65), followed by the standard errors of $\widehat{\mathsf{B}}_{\mathrm{dCov}}$ and the remaining two procedures. Higher standard errors (resulting in larger confidence intervals in Figure 2) might arise from non-unique identification with respect to signs and ordering of the columns of the structural matrix $\widehat{\mathsf{B}}$. In particular, CvM and dCov identification involves the multiplication of B with rotation matrices so that the column order and signs potentially have to be adjusted after identification. In contrast, during ML identification the ordering remains unchanged neglecting alternative permutations.

Next, we study the dynamic impact of the identified structural shocks on oil production, real economic activity and the real price of oil. Figure 1 displays the impulse response functions (IRFs) calculated from the estimated reduced form model and the identified matrices in (8). The IRFs appear quite similar in shape to those presented in Kilian and Murphy (2012) and Lütkepohl and Netšunajev (2014). In most cases, we can easily distinguish between the IRFs based on ML and dCov identification, which nearly coincide, and those based on the CvM statistic and sign restrictions. Furthermore, the simulation results, pointing to biased matrix estimates for the sign restriction approach, are corroborated by the responses to an oil supply shock displayed in the first column. As eyeballing the IRFs suggests, this bias might have caused a shift in the impulse responses. Related to the non-uniqueness of column ordering and signs in the case of independence based identification, it seems more challenging to distinguish between the responses to an aggregate and an oil-specific demand shock. The IRFs from ML and dCov identification are not in line with the economic intuition (Table 5). Nevertheless, they support the results of Kilian and Murphy (2012) and Lütkepohl and Netšunajev (2014), even more pronounced than suggested in Herwartz and Plödt (2016b). Especially for the oil-specific demand shock, these IRFs based on additional economic assumptions (Kilian and Murphy, 2012) appear very similar.

We further examine the response of the real price of oil to an oil supply, an aggregate demand and an oil-specific demand shock in Figure 2 (in line with the study of Kilian and Murphy (2012)). The displayed dynamic profiles of the IRFs and the corresponding 95%

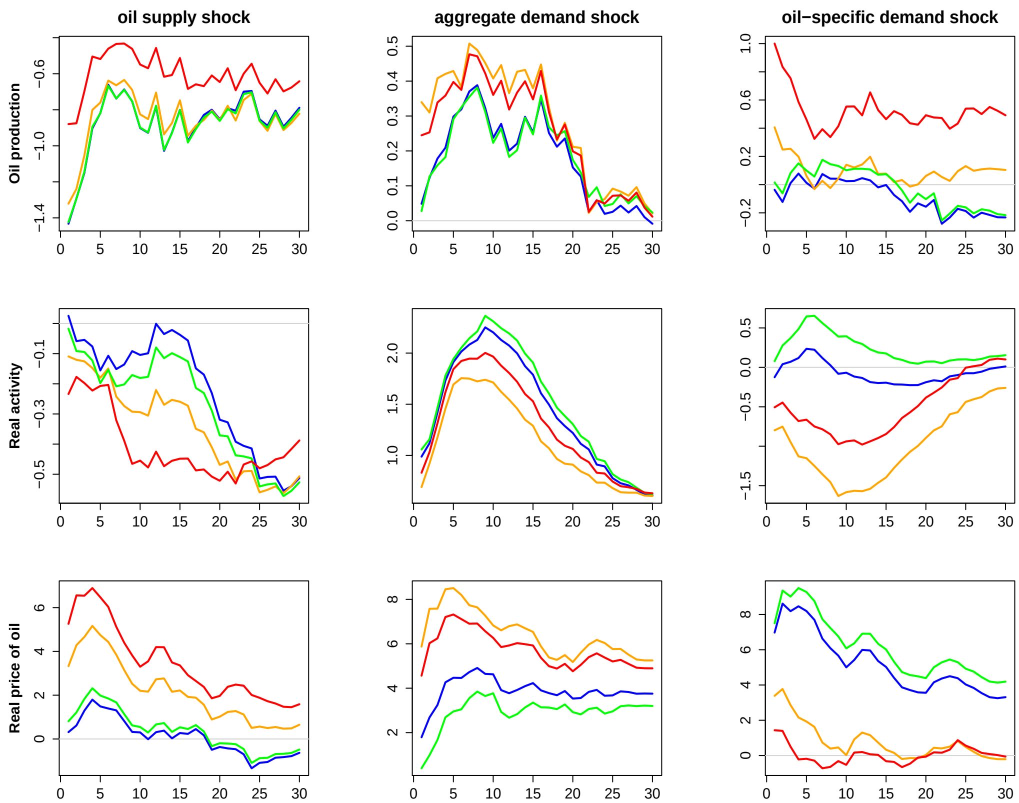


Figure 1: Impulse response functions based on OLS estimation of the VAR(24) model in (1) and identification based on sign restrictions (median of successful draws, red), the ML method (blue), dCov (green) and CvM (orange).

confidence intervals further underpin the conclusions drawn above. While exhibiting higher uncertainty on impact, the IRFs of the identification approach employing dCov as dependence measure closely resemble those of the ML method. Nevertheless, it might be noteworthy that at a horizon of twenty months the confidence intervals of the alternative identification schemes appear similar in size. CvM identification generates impulse responses rather in line with those obtained by sign restrictions. However, the corresponding confidence intervals include the IRFs obtained from the alternative independence based identifications. Again, for the CvM identification scheme wider confidence intervals might arise from the difficulty to determine the most appropriate column order.

5 Conclusions

In SVAR models uncorrelated structural shocks are commonly identified by means of economically motivated restrictions or by statistical means. In this sense we focus on a classical version of identification based on sign restrictions (representing economic restrictions) and three identification procedures based on non-Gaussianity of the shocks (statistical identification). More precisely, the ML method of Lanne et al. (2017) and two dependence diagnostics, namely the criteria applied by Herwartz (2015) and Matteson and Tsay (2013), promise consistent estimation of uniquely identified independent structural shocks. We compare the alternative identification approaches in a simulation study and an application to the global crude oil market.

By means of Monte Carlo simulations we confirm and specify the bias induced by classical sign restrictions to occur irrespective of the underlying distribution and sample size. In contrast, independence based identification provides (consistent) matrix estimates with decreasing MSE for increasing sample size. Accordingly, the frequency of correct sign patterns (i.e., the sign pattern of the data generating structural matrix) is higher for these estimated in larger samples. Furthermore, we can differentiate between the parametric identification procedure based on maximizing the Student-t likelihood and the two nonparametric dependence measures. For a correctly specified likelihood and sufficiently large sample sizes the ML procedure performs best. The nonparametric dependence diagnostics benefit if the distribution deviates from the one assumed a-priori. Moreover, in identifying χ^2-distributed or t-distributed structural shocks their performance is markedly better for the former.

Additionally, we apply the four identification procedures to an SVAR model of the global crude oil market. As a result, supply, aggregate demand and oil-specific demand shocks can be distinguished. The alternative estimates quantifying instantaneous transmissions from structural shocks to observables and the corresponding impulse response functions appear in line with the findings from the simulation study. At the same time, the impulse responses support the results of Kilian and Murphy (2012) stating that stylized sign restrictions might not be sufficient for identification and lead to biased estimation. While Kilian and Murphy (2012) argue for further theory based assumptions, the independence based approaches do not restrict a-priori economically reasonable impulse responses. Nevertheless, providing a similar dynamic pattern, independence based identification might be less restrictive, especially, if the theoretical background is a subject of discussion.

References

Arias, J., Rubio-Ramírez, J., Waggoner, D., 2014. Inference based on svars identified with sign and zero restrictions: Theory and applications. Dynare Working Paper Series 30.

Blanchard, O. J., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. American Economic Review 79 (4), 655–73.

Canova, F., Nicolo, G. D., 2002. Monetary disturbances matter for business fluctuations in the g-7. Journal of Monetary Economics 49 (6), 1131–1159.

Capasso, M., Moneta, A., 2016. Macroeconomic responses to an independent monetary policy shock: a (more) agnostic identification procedure. Lem papers series, Sant'Anna School of Advanced Studies, Pisa, Italy.

Carlstrom, C. T., Fuerst, T. S., Paustian, M., October 2009. Monetary policy shocks, Choleski identification, and DNK models. Journal of Monetary Economics 56 (7), 1014–1021.

Castelnuovo, E., Oct. 2012. Monetary policy neutrality: Sign restrictions go to Monte Carlo. "Marco Fanno" Working Papers 151, Dipartimento di Scienze Economiche "Marco Fanno".

Castelnuovo, E., 2013. Monetary policy shocks and financial conditions: A Monte Carlo experiment. Journal of International Money and Finance 32, 282–303.

Castelnuovo, E., 2016. Monetary policy shocks and cholesky vars: An assessment for the euro area. Empirical Economics 50 (2), 383–414.

Chib, S., Ramamurthy, S., 2014. Dsge models with student-t errors. Econometric Reviews 33 (1-4), 152–171.

Comon, P., 1994. Independent component analysis, A new concept? Signal Processing 36 (3), 287–314.

Cúrdia, V., Del Negro, M., Greenwald, D. L., 2014. Rare shocks, great recessions. Journal of Applied Econometrics 29 (7), 1031–1052.

Faust, J., 1998. The robustness of identified var conclusions about money. Carnegie-Rochester Conference Series on Public Policy 49 (1), 207–244.

Fry, R., Pagan, A., 2007. Some issues in using sign restrictions for identifying structural vars. NCER Working Paper Series 14, National Centre for Econometric Research.

Fry, R., Pagan, A., 2011. Sign restrictions in structural vector autoregressions: A critical review. Journal of Economic Literature 49 (4), 938–60.

Genest, C., Rémillard, B., 2004. Tests of independence and randomness based on the empirical copula process. Test 13 (2), 335–370.

Gertler, M., Gali, J., Clarida, R., 1999. The science of monetary policy: A new Keynesian perspective. Journal of Economic Literature 37 (4), 1661–1707.

Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89–120.

Gouriéroux, C., Monfort, A., Renne, J.-P., 2017. Statistical inference for independent component analysis: Application to structural var models. Journal of Econometrics 196, 111–126.

Herwartz, H., 2015. Structural var modelling with independent innovations - an analysis of macroeconomic dynamics in the euro area based on a novel identification scheme, mimeo.

Herwartz, H., Lütkepohl, H., 2014. Structural vector autoregressions with Markov switching: combining conventional with statistical identification of shocks. J. Econometrics 183 (1), 104–116.

Herwartz, H., Maxand, S., 2017. Nonparametric tests for independence - a review and comparative simulation study with an application to malnutrition data in india, mimeo.

Herwartz, H., Plödt, M., 2016a. Simulation Evidence on Theory-based and Statistical Identification under Volatility Breaks. Oxford Bulletin of Economics and Statistics 78, 94–112.

Herwartz, H., Plödt, M., 2016b. The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. Journal of International Money and Finance 61, 30–44.

Hodges, J., Lehmann, E., 2006. Hodges-lehmann estimators. In: Encyclopedia of Statistical Sciences.

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2015. copula: Multivariate Dependence with Copulas. R package version 0.999-13.

Hyvärinen, A., Zhang, K., Shimizu, S., Hoyer, P. O., 2010. Estimation of a structural vector autoregression model using non-Gaussianity. J. Mach. Learn. Res. 11, 1709–1731.

Kilian, L., 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. American Economic Review 99 (3), 1053–69.

Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press, forthcoming.

Kilian, L., Murphy, D. P., 2012. Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models. Journal of the European Economic Association 10 (5), 1166–1188.

Lanne, M., Lütkepohl, H., 2008. Identifying monetary policy shocks via changes in volatility. Journal of Money, Credit and Banking 40 (09), 1131–1149.

Lanne, M., Lütkepohl, H., 2010. Structural vector autoregressions with nonnormal residuals. J. Bus. Econom. Statist. 28 (1), 159–168.

Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. J. Econometrics 196 (2), 288–304.

Lewbel, A., 2010. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. Journal of Business & Economic Statistics 30 (1), 67–80.

Lütkepohl, H., Netšunajev, A., 2014. Disentangling demand and supply shocks in the crude oil market: how to check sign restrictions in structural VARs. J. Appl. Econometrics 29 (3), 479–496.

Matteson, D. S., Tsay, R. S., 2013. Independent component analysis via distance covariance. pre-printHttp://arxiv.org/abs/1306.4911.

Moneta, A., Entner, D., Hoyer, P. O., Coad, A., 2013. Causal inference by independent component analysis: Theory and applications. Oxford Bulletin of Economics and Statistics 75 (5), 705–730.

Newey, W. K., Steigerwald, D. G., 1997. Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. Econometrica 65 (3), 587–599.

Peersman, G., Van Robays, I., 2012. Cross-country differences in the effects of oil shocks. Energy Economics 34 (5), 1532–1547.

Rigobon, R., 2003. Identification through heteroskedasticity. The Review of Economics and Statistics 85, 777–792.

Risk, B. B., James, N. A., Matteson, D. S., 2015. steadyICA: ICA and Tests of Independence via Multivariate Distance Covariance. R package version 1.0.

Sims, C. A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.

Székely, G. J., Rizzo, M. L., Bakirov, N. K., 2007. Measuring and testing dependence by correlation of distances. Ann. Statist. 35 (6), 2769–2794.

Uhlig, H., 2005. What are the effects of monetary policy on output? results from an agnostic identification procedure. Journal of Monetary Economics 52 (2), 381–419.

Appendix

The 3-equation DSGE model

For simulation purposes we employ a simple 3-equation dynamic stochastic general equilibrium (DSGE) model that has been widely used as a baseline framework for monetary policy analysis (Gertler et al., 1999; Carlstrom et al., 2009; Castelnuovo, 2013, 2012, 2016). The consideration of trivariate systems is also common practice in the SVAR literature. The log-linearized version of the model reads as

$$\begin{aligned} x_t &= \gamma E_t x_{t+1} + (1-\gamma)x_{t-1} - \delta_x(r_t - E_t\pi_{t+1}) + \omega_{x,t}, && (9)\\ \pi_t &= (1+\alpha\beta)^{-1}\beta E_t\pi_{t+1} + (1+\alpha\beta)^{-1}\alpha\pi_{t-1} + \kappa x_t + \omega_{\pi,t}, && (10)\\ r_t &= \tau_r r_{t-1} + (1-\tau_r)(\tau_\pi\pi_t + \tau_x x_t) + \omega_{r,t}, && (11)\\ \omega_{\bullet,t} &= \rho_\bullet\omega_{\bullet,t-1} + \varepsilon_{\bullet,t}, \bullet \in \{x,\pi,r\},\ t=1,\ldots,T, && (12) \end{aligned}$$

where x_t, π_t and r_t denote the output gap, inflation and the nominal interest rate, respectively, and E_t indicates expectations formed at period t. Accordingly, the equations (9) to (11) represent a New Keynesian IS equation, a hybrid New Keynesian Phillips curve, and a Taylor rule with interest rate smoothing. First order autoregressive shock processes are summarized in (12), with subscripts $\bullet \in \{x, \pi, r\}$ indicating a demand shock, a supply shock and a monetary policy shock, respectively.

The employed parameter settings correspond to common calibration assumptions drawn from the macroeconomic literature. The model is calibrated with common settings, i.e., $\beta = 0.99$ (discounting), $\kappa = 0.05$ (slope of Phillips curve), $\alpha = 0.5$ (indexation of past inflation), $\delta_x = 0.1$ (impact of real interest), $\gamma = 0.5$ (effect of output expectations), $\tau_\pi = 1.8, \tau_x = 0.5, \tau_r = 0.6$ (Taylor rule). The autoregressive parameters in (12) are set to $\rho_x = \rho_\pi = \rho_r = 0.5$.

Dependence Diagnostics

Cramér-von Mises statistic (Genest and Rémillard, 2004)

Mutual dependence within a K-dimensional vector of structural shocks ε_t at time $t = 1, \ldots, T$ can be measured by the Cramér-von Mises functional

$$\mathcal{C} = \int_{(0,1)^K} T\Big(C_T(\epsilon) - \prod_{k=1}^{K} U_T(\epsilon_k)\Big)^2 d\epsilon$$

with cumulative distribution function U_T of a uniformly distributed variable on $\{1/T, \ldots, T/T\}$ and the empirical copula C_T. Apparently, the functional $\mathcal{C}$ measures the distance between the empirical copula based on the vector of structural shocks ε_t and the copula under independence. Genest and Rémillard (2004) describe the estimation of the copula and the explicit statistic in more detail. Minimizing $\mathcal{C}$ with respect to B (i.e., considering an empirical copula C_T determined by $\hat{\varepsilon}_t = \mathsf{B}^{-1}\hat{u}_t$) provides the HL estimates and the corresponding least dependent components.

Distance covariance (Matteson and Tsay, 2013)

For a K-dimensional vector of structural shocks ε_t at time $t = 1, \ldots, T$ the distance covariance $\mathcal{V}^2$ detects dependence between two subsets of the components. Between the kth component

$\varepsilon_{t,k}$, $k \in \{1, \ldots, K\}$ and all subsequent ones ε_{t,k^+} with $k^+ = k+1, \ldots, K$, dependence is measured by $\mathcal{V}^2(\varepsilon_{t,k}, \varepsilon_{t,k^+})$ which is the distance between the characteristic functions $\varphi_{\varepsilon_{t,k},\varepsilon_{t,k^+}}$ and $\varphi_{\varepsilon_{t,k}}\varphi_{\varepsilon_{t,k^+}}$, the joint characteristic function and the one under independence, respectively. To measure mutual dependence, i.e. dependence of all possible combinations between the variables $\varepsilon_{t,1}, \ldots, \varepsilon_{t,K}$, the dependence criterion reads as

$$\mathcal{U}_T(\varepsilon_{t,1}, \ldots, \varepsilon_{t,K}) = T \cdot \sum_{k=1}^{K-1} \mathcal{V}^2(\varepsilon_{t,k}, \varepsilon_{t,k^+}). \tag{13}$$

The distance covariance $\mathcal{U}_T(\hat{\varepsilon}_{t,1}, \ldots, \hat{\varepsilon}_{t,K})$ is then minimized to identify $\hat{\varepsilon}_t = \mathsf{B}^{-1}\hat{u}_t$ with least dependent components.

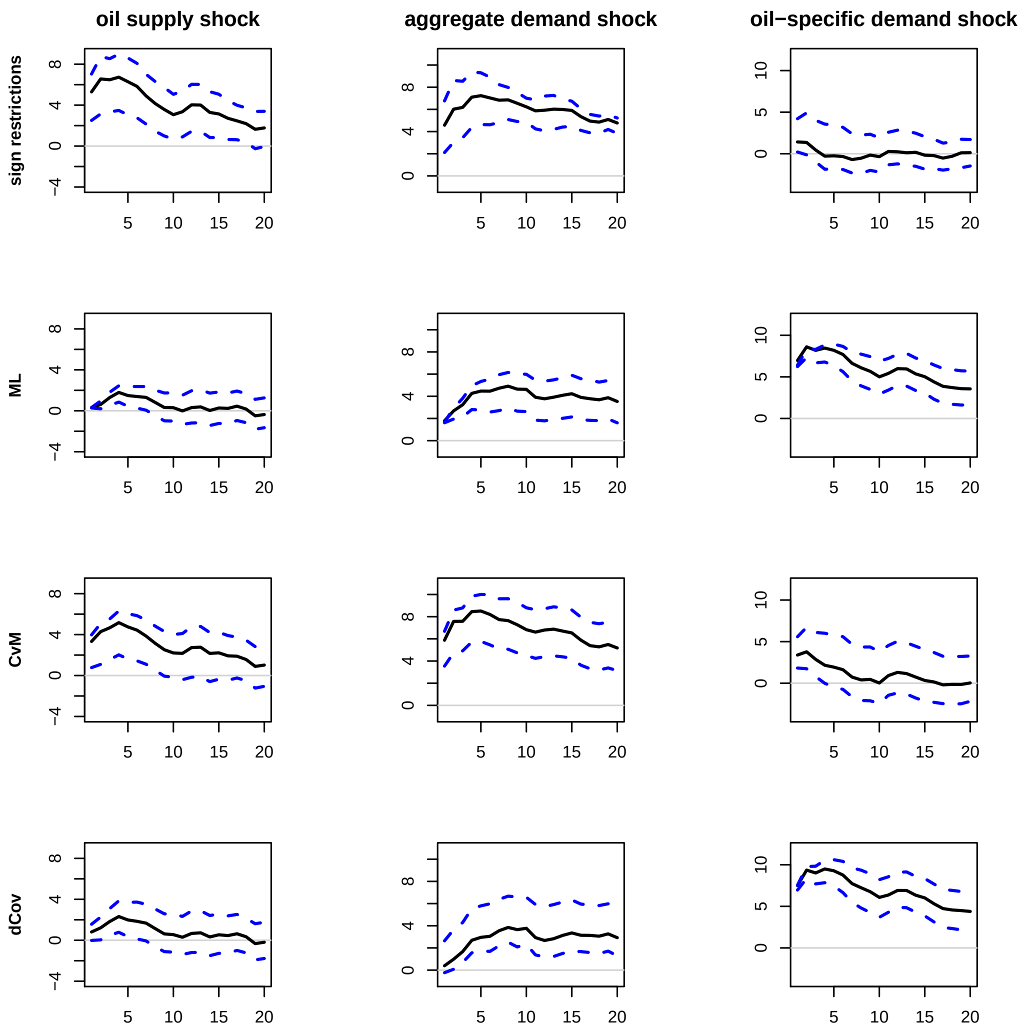


Figure 2: Response of the real price of oil based on OLS estimation of the VAR(24) model in (1) joint with 95% confidence intervals.

D Identification of independent structural shocks in the presence of multiple Gaussian components

Identification of independent structural shocks in the presence of multiple Gaussian components

Simone Maxand*

University of Goettingen

September 27, 2017

Abstract

Several recently developed identification techniques for structural VAR models are based on the assumption of non-Gaussianity. So-called independence based identification provides unique structural shocks (up to scaling and ordering) under the assumption of at most one Gaussian component. While non-Gaussianity of certain interesting shocks, e.g., a monetary policy shock, appears rather natural, not all macroeconomic shocks in the system might show this clear difference from Gaussianity. We generalize identifiability by noting that even in the presence of multiple Gaussian shocks the non-Gaussian ones are still unique. Consequently, independence based identification allows to uniquely determine the (non-Gaussian) shocks of interest irrespective of the distribution of the remaining system. In an illustrative macroeconomic model the identified structural shocks confirm the results of previous studies on the early millennium slowdown. Furthermore, extending the time horizon provides full identification under the non-Gaussianity assumption.

Keywords: SVAR, identification, non-Gaussian, millennium slowdown.

JEL Classification: C32, E32.

*Chair of Econometrics, University of Goettingen, Humboldtallee 3, D-37073 Goettingen, Germany. Telephone: +49 551 3924604, e-mail: *smaxand@uni-goettingen.de*. I thank Raffaella Giacomini and Hannes Rohloff for helpful comments and suggestions.

1 Introduction

Structural vector autoregressive (SVAR) models are frequently applied to identify the fundamental economic driving forces in macroeconomic systems. In this framework, diverse approaches aim at tracing macroeconomic variables back to orthogonal shocks (see Kilian and Lütkepohl, 2017, for an overview). While the identification procedures handle non-uniqueness of the structural matrix by building on certain statistical or economic assumptions, the views on the adequacy of these restrictions are diverging. Under Gaussianity, additional economic restrictions help to reduce the set of uncorrelated structural shocks, derived by any decomposition of the covariance matrix, to those in line with common economic beliefs (Sims, 1980; Blanchard and Quah, 1989; Faust, 1998; Uhlig, 2005). However, uncorrelated non-Gaussian structural shocks can still incorporate diverse forms of dependence. In order to separate the shocks and the associated responses completely, independent component analysis (ICA) methods uniquely identify the instantaneous response matrix for independent structural shocks under non-Gaussianity. These approaches base on the prominent theorem of Comon (1994) which indicates the existence of a unique structural matrix if the model contains at most one Gaussian structural shock (see, for instance, Moneta et al., 2013; Gouriéroux et al., 2017; Lanne et al., 2017).

When applying a structural VAR model the analyst is mostly interested in studying the responses to certain shocks only. For instance, the macroeconomic implications of monetary policy shocks have been widely analyzed by means of SVAR techniques. The distribution of the change in interest rates, estimated by a kernel density in Figure 1 (cf. Chiu et al., 2016), leads to the rather natural assumption that an unanticipated shock in monetary policy comes from a non-Gaussian distribution. However, different macroeconomic variables might be more 'balanced' in that they follow a distribution which is closer to Gaussianity (e.g. a supply or demand shock). In order to identify only parts of the system, we allow the K-dimensional vector of structural shocks ε_t to contain $1 < k_1 < K$ Gaussian components. In this setting, neither Gaussianity implies independence of all shocks nor ICA methods can just identify the whole system. We show that the $K - k_1$ non-Gaussian components of ε_t can still be uniquely identified by ICA methods. This result introduces flexibility by allowing for partial identification of the system after diagnosing (non-) Gaussianity of the structural shocks. Especially, when the effect of only certain structural shocks is of interest (and they are non-Gaussian), the distribution of the remaining system is irrelevant for their identification.

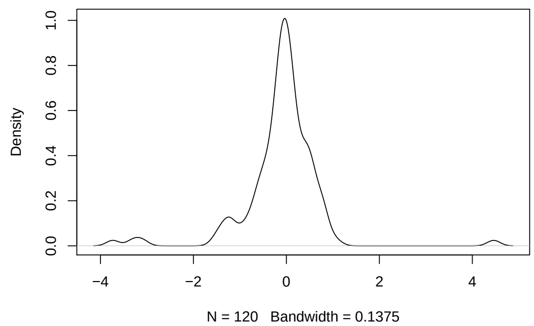


Figure 1: Kernel density estimate of the change in nominal interest rate in 1984–2002 (for a more detailed description of the data see Section 3).

We illustrate partial identification by re-investigating a four dimensional macroeconomic model in the spirit of Peersman (2005) who intended to identify the causes of the early millennium slowdown. More specifically, we identify two of four possible independent shocks by relying on a nonparametric dependence measure, the distance covariance. Studying quarterly data for 1980–2002, we interpret the identified oil price and monetary policy shocks in light of former replication studies. For an extended sample, more pronounced differences from Gaussianity arise. This allows full identification of the system and the interpretation of the response to all structural shocks.

In Section 2, we describe the model setting and the identification techniques for at most one and multiple Gaussian components. Section 3 contains the description and discussion of the estimation results for a four dimensional macroeconomic model. Section 4 concludes.

2 Model and identification

We consider a $K-$dimensional macroeconomic VAR model formulated as

$$\begin{aligned} y_t &= c_t + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \\ &= c_t + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \mathsf{B}\varepsilon_t = \mu + \sum_{i=0}^{\infty} \Phi_i \mathsf{B}\varepsilon_{t-i} \quad t = 1, \ldots, T, \end{aligned} \tag{1}$$

where c_t is a matrix of deterministic terms, y_t is $K \times 1$ dimensional and $A_1, \ldots, A_p$ and Φ_i are $K \times K$ matrices. For paraphrasing (1) we assume causality of the model, i.e., $\det \Phi(z) \neq 0$ for all $|z| \leq 1$ with $\Phi(z) = \sum_{i=0}^{\infty} \Phi_i z^i$ and $\Phi_0 = I_K$. Reduced form residuals correspond to error terms $u_t \sim (0, \Sigma_u)$ with non-singular covariance matrix $\Sigma_u = \mathsf{B}\mathsf{B}'$. The main interest of the following study is the identification of matrix B and the associated structural shocks $\varepsilon_t = \mathsf{B}^{-1}u_t$ with $E(\varepsilon_t) = 0$ and $\Sigma_\varepsilon = \mathsf{B}^{-1}\Sigma_u\mathsf{B}^{-1} = I_K$. For this purpose, the literature on SVAR models incorporates numerous approaches to identify the non-unique factor B properly relying on either statistical or economic a-priori assumptions (for a textbook treatment of SVARs see Kilian and Lütkepohl, 2017).

2.1 Independence based identification

Recently developed statistical identification procedures exploit the non-normality of structural shocks building on results from independent component analysis (Moneta et al., 2013; Lanne et al., 2017; Gouriéroux et al., 2017). For the vector of reduced form errors $u_t \in \mathbb{R}^K$, ICA aims at determining the so-called mixing matrix B for which the components of $\mathsf{B}^{-1}u_t = \varepsilon_t$ are independent. Following the fundamental result of Comon (1994), ICA uniquely identifies matrix B up to column signs and ordering by allowing the vector of independently distributed structural shocks ε_t to contain at most one Gaussian component $\varepsilon_{t,k}$.

In the following, we describe identification in the case of one and multiple Gaussian components on the basis of an ICA procedure adapted from Matteson and Tsay (2017). The distance covariance, a nonparametric dependence measure introduced in Székely et al. (2007), is applied to determine least dependent shocks and thereby, to identify the associated matrix B. It might be noteworthy that similar ICA-based identification procedures lead to the same theoretical results in the case of multiple Gaussian components.

2.1.1 Identification with at most one Gaussian structural shock

Moneta et al. (2013) have adopted ICA to determine optimal variable orderings in recursive systems of non-Gaussian structural shocks. However, the a-priori focus on triangular schemes appears restrictive in an economic context. Determining the underlying distribution family a-priori, Lanne et al. (2017) apply ML estimation to determine the matrix B. Moreover, nonparametric dependence measures provide an alternative tool for identification avoiding any restrictive assumption on the distribution of ε_t. In this work, we rely on the so-called distance covariance of Székely et al. (2007) applied in the course of ICA by Matteson and Tsay (2017).[1] The set of possible decompositions

[1] Diverse alternative criteria have been studied in preliminary analyses (avalaible on request) where especially the Cramér-von Mises distance turns out as a robust alternative to measure dependence nonparametrically.

of the least squares covariance estimator $\mathsf{B}(\theta) = DQ(\theta)$ is defined with respect to Choleski factor D and the vector of rotation angles θ of the Givens matrices $Q(\theta)$. We estimate the covariance matrix once by least squares and different decompositions evolve by drawing from the set of all rotation angles θ. Accordingly, the distance covariance $\mathcal{U}_T(\hat{\varepsilon}_t(\theta))$ can be calculated from $\hat{\varepsilon}_t(\theta) = \mathsf{B}(\theta)^{-1}\hat{u}_t$ where $\hat{u}_t$ are the least squares residuals. Minimization of the distance covariance $\hat{\theta} = \text{argmin}_\theta \mathcal{U}_T(\hat{\varepsilon}_t(\theta))$ consequently determines the estimated matrix $\widehat{\mathsf{B}} = \mathsf{B}(\hat{\theta})$ and the associated least dependent shocks $\hat{\varepsilon}_t(\hat{\theta})$. For details on the exact minimization procedure and the empirical definition of the dependence measure we refer to Matteson and Tsay (2017). In this study, we apply the function *steadyICA* from the `R` package `steadyICA` (Risk et al., 2015) to determine $Q(\hat{\theta})$ and thus, $\widehat{\mathsf{B}}_{\text{dCov}} = \mathsf{B}(\hat{\theta})$.

2.1.2 Identification with multiple Gaussian structural shocks

More generally, let the vector ε_t contain $1 \leq k_1 \leq K$ Gaussian random variables. If the number of Gaussian components exceeds one, i.e. $k_1 > 1$, matrix B can no longer be uniquely identified and consequently, the structural shocks $\varepsilon_t = \mathsf{B}^{-1}u_t$ can not be separated by means of ICA. However, by an intuitive generalization of Comon's theorem the $K - k_1$ non-Gaussian components of ε_t remain unique. We formulate this result in the following proposition for two random vectors $\varepsilon_1, \varepsilon_2 \in \mathbb{R}^K$, representative for vectors with independent components not distinguishable by means of ICA. Within these vectors the Gaussian components are ordered first.

Proposition 1. *Let ε_1 be a vector with independent components of which only w.l.o.g. the first k_1 components are Gaussian. Let C be an orthogonal $K \times K$ matrix and $\varepsilon_2 = C\varepsilon_1$ such that the first k_1 entries of ε_2 are Gaussian. The components of ε_2 are mutually independent if and only if $C = \begin{pmatrix} Q & 0 \\ 0 & \Lambda P \end{pmatrix}$ where matrix Q is an orthogonal $k_1 \times k_1$ matrix, Λ is a $(K - k_1) \times (K - k_1)$ diagonal matrix and P is a permutation matrix.*

The proof is given in the Appendix and represents an alternative to Boscolo et al. (2002). For matrix C as defined in Proposition 1, ICA can not distinguish between $\mathsf{B}C$ and B, in other words $\varepsilon_t = (\mathsf{B}C)^{-1}u_t$ also comprises independent components. In the following, we apply the ICA procedure of Matteson and Tsay (2017) to models with several Gaussian structural shocks. Statistical properties, as consistency, of the *steadyICA* algorithm under multiple Gaussian components transfer to the subsample of non-Gaussian variables. Leaving the formal derivation aside we assume that the first k_1 columns of $\widehat{\mathsf{B}}_{\text{dCov}}$ (if Gaussian components are ordered first) are not uniquely determined as the Gaussian variables can not be distinguished (Hyvärinen et al., 2001). In contrast, the remaining $K - k_1$ columns of $\widehat{\mathsf{B}}_{\text{dCov}}$ are unique. Along these lines, for at most one Gaussian component all columns of $\widehat{\mathsf{B}}_{\text{dCov}}$ are unique. For applicability of the identification technique it is essential to decide on the number of Gaussian components first.

Decide on the number of Gaussian components

Various alternative uni- and multivariate tests for normality are present in the literature. A selection of tests is, for instance, implemented in the `R` package `normtest` (Gavrilov and Pusev, 2015). Moreover, diverse strategies can be pursued to assess normality of a multivariate vector of structural shocks $\hat{\varepsilon}_t$. In the following, we choose two alternative approaches. First, we test separately on Gaussianity of the components and secondly, we apply a test which decides on the number of non-Gaussian components in ICA. The results of separate univariate Jarque-Bera (JB) tests provide evidence for Gaussianity of the structural shocks determined by independence based identification, e.g. $\hat{\varepsilon}_t = \widehat{\mathsf{B}}_{\text{dCov}}^{-1}\hat{u}_t$. Note that the results from alternative univariate tests provide similar test outcomes and are not displayed here. Under the null hypothesis of the JB test the shock exhibits a Gaussian distribution. Thus, if the null hypothesis is rejected we assume that the associated shock can be uniquely identified by means of ICA.

However, the estimated structural shocks $\hat{\varepsilon}_t$ and their distribution might depend on the underlying identification procedure. To evaluate robustness of the JB test decisions, we apply techniques based on fourth order blind identification (FOBI) which have evolved in the course of non-Gaussian component analysis (NGCA) to isolate non-Gaussian from Gaussian components. In their `R` package *ICtest*, Nordhausen et al. (2016) have implemented several tests to decide on the number of non-Gaussian, so-called interesting, components within a set of variables. We apply the version implemented in the function `FOBIboot` which uses a bootstrap procedure. The test applies FOBI to trace the vector of reduced form residuals back to Gaussian and non-Gaussian sources. The corresponding null hypothesis states that there are k_1 Gaussian components and $K - k_1$ non-Gaussian components. For further details on the test and the implementation we refer to the manual of the `R` package (Nordhausen et al., 2016).

It might be noteworthy that the JB tests on Gaussianity of the structural shocks and the application of one overall test for Gaussian components provide a test decision derived under different significance levels. Either four separate tests on a certain level are performed or one single test helps, for instance, to decide about two Gaussian components on one level. We apply and compare both approaches in the subsequent application to a four dimensional macroeconomic model.

3 Reassessing causes of the early millennium slowdown

We consider the model in (1) where now $y_t = (\Delta oil_t, \Delta y_t, \Delta p_t, s_t)$ contains first differences of oil prices Δoil_t, output growth Δy_t, consumer inflation Δp_t and the short term interest rate s_t. Peersman (2005) applies this model setting to study the causes of the early millennium slowdown in 2001. In the following, we will consider the model in two variations of the sample period. First, we replicate the study of Peersman (2005) for the original sample 1980Q1–2002Q2. An extended sample includes data until 2007Q4 to further assess causes of the slowdown in 2001.[2] For the two samples we examine applicability of independence based identification by assessing Gaussianity of the shocks. Furthermore, we analyze the impulse responses estimated by means of the technique which relies on the distance covariance.

	$\hat{\varepsilon}_1$	$\hat{\varepsilon}_2$	$\hat{\varepsilon}_3$	$\hat{\varepsilon}_4$
JB	56.225	1.045	0.060	23.686
p-value	0.000	0.527	0.969	0.005

H_0 :	$k_1 = 2$	$k_1 = 3$
Test Stat.	16.312	491.88
p-value	0.915	0.035

Table 1: JB test results for $\hat{\varepsilon}_t = \widehat{\mathsf{B}}_{\text{dCov}}^{-1}\hat{u}_t$ for sample 1980Q1–2002Q2 (left-hand side table). Tests on non-Gaussian components in $\hat{u}_t$: we can reject that there are $k_1 = 3$ Gaussian components but we can not reject that there are $k_1 = 2$ Gaussian components at a reasonable significance level.

Table 1 and 2 display the outcome of separate JB tests for the structural shocks $\hat{\varepsilon}_t = \widehat{\mathsf{B}}_{\text{dCov}}^{-1}\hat{u}_t$ and sample periods 1980Q1–2002Q2 and 1980Q1–2007Q4, respectively.[3] Alongside, we display statistics and p-values of the tests on interesting, i.e. non-Gaussian, components. The JB test results hint at the presence of two Gaussian components ε_2 and ε_3 on the shorter horizon (Table 1). In the larger sample we reject normality of three of the four components at 10% significance level based on the JB tests (Table 2). By means of the test on interesting components we obtain

[2] It might be noteworthy that Peersman (2005) studies data for the US, the Euro area and the industrialized world. He argues that the effects appear the most pronounced in the US. As noted by Grant (2015) differences between the results and Peersman (2005) may occur due to data deviations.

[3] Note that slight differences to exact p-values might be caused by the Monte Carlo simulation used for calculation of the test distribution. However, we assume that the main conclusions remain unchanged.

the same result in the smaller sample. However, relying on this test, we might still assume the presence of two Gaussian components in the larger sample.

	$\hat{\varepsilon}_1$	$\hat{\varepsilon}_2$	$\hat{\varepsilon}_3$	$\hat{\varepsilon}_4$
JB	63.272	3.623	0.476	70.257
p-value	0.000	0.099	0.76	0.000

H_0 :	$k_1 = 2$	$k_1 = 3$
Test Stat.	69.288	1369
p-value	0.602	0.005

Table 2: JB test results for $\hat{\varepsilon}_t = \widehat{\mathsf{B}}_{\text{dCov}}^{-1}\hat{u}_t$ for sample 1980Q1–2007Q4. Tests on non-Gaussian components in $\hat{u}_t$: we can reject that there are $k_1 = 3$ Gaussian components but we can not reject that there are $k_1 = 2$ Gaussian components at 10% significance level.

Following Section 2.1.2, we assume that the distance covariance uniquely identifies the non-Gaussian shocks in the smaller and all shocks in the larger sample (relying on the JB test results). Further differences caused by the sample choice are reflected in the impulse responses in Figure 2 calculated using independence based identification. The displayed confidence intervals are calculated from a wild bootstrap procedure as, for instance, described in Herwartz and Plödt (2016). First, we notice that the confidence intervals in the shorter sample are mostly wider. This seems an intuitive consequence of the larger and more likely identified (because of non-Gaussianity) model exhibiting smaller estimation uncertainty. Furthermore, the point estimates of the dynamic responses are partly shifted which we attribute to a change in the data (i.e. the relations between variables) as well as the adequacy of the identification approach. However, in both cases we obtain two uniquely identified shocks, the first and the fourth, and can observe that the corresponding impulse responses appear very similar in both samples. Based on the reasoning of the following paragraph we label the first shock an oil price and the fourth a monetary policy shock. For these derivations we proceed with the model including data up to 2007Q4, merely to overcome the identification issues. However, it might be noteworthy that the results for the oil price and the monetary policy shock hold similar in the smaller sample period.

In order to label the shocks adequately based on Figure 2, we rely on former replication studies by Herwartz and Lütkepohl (2014) and Lanne and Luoto (2016). In the last column of Figure 2 we almost exactly replicate the responses to a monetary policy shock obtained by the method of Herwartz and Lütkepohl (2014). Also in line with the results of Uhlig (2005) and Lanne and Luoto (2016), the sign pattern suggested in Peersman (2005) is thereby not replicated. Furthermore,

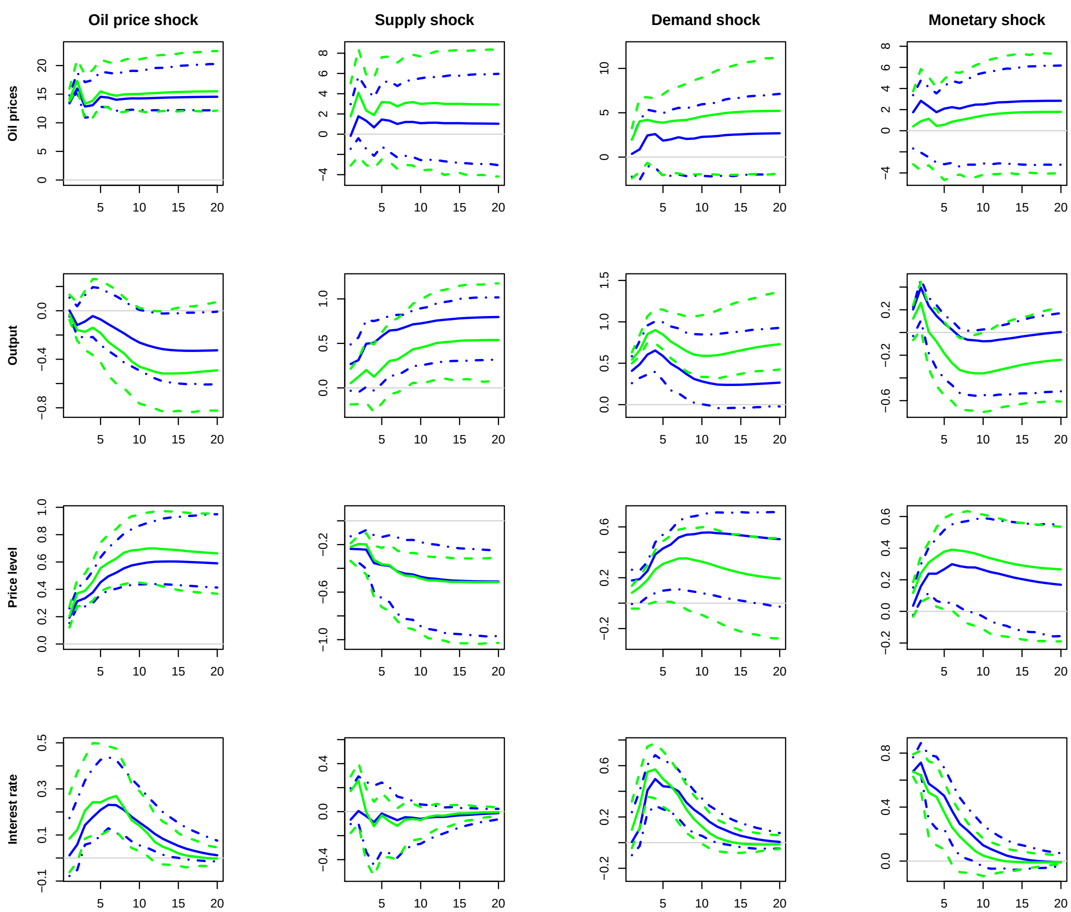


Figure 2: Impulse response functions based on identification by means of distance covariance for samples 1980Q1-2002Q2 (green, dashed confidence intervals) and 1980Q1-2007Q4 (blue, dotdashed confidence intervals).

Lanne and Luoto (2016) argue that only the oil price shock can be fully reproduced holding the suggested signs in the on-impact matrix. Acknowledging higher uncertainty in the instantaneous responses, we therefore label the first shock an oil price shock. The supply and demand shock both lead to insignificant responses in the associated variables and thus, might not be identifiable. However, the assigned labels appear economically reasonable and further support the results of Herwartz and Lütkepohl (2014) and Lanne and Luoto (2016). Overall, the impulse responses displayed in Figure 2 still indicate that a combination of shocks causes the slowdown in the short as well as in the long run. However, output does not seem to respond significantly to a monetary policy shock.

Decomposing output growth into the contribution of structural shocks in each time period provides further evidence on the causes of negative economic growth in 2001. Figure 3 shows the corresponding historical decompositions starting in 1995 up to 2007 (calculated as described in Lütkepohl, 2011). Based on Figure 3, the recession in 2001 is attributed to a combination of shocks which is in line with the conclusions drawn in Peersman (2005). Yet the size and direction of the contributions vary throughout the time periods of output declines. While in the third quarter of 2001 all shocks dampen output growth with roughly the same impact, their contributions in early 2001 differs. The aggregate demand shock provokes the largest negative contribution in quarter 1 of 2001 which is subsequently slightly positive in quarters 2 and 4. Throughout 2001 monetary policy further reduces output growth while the contribution becomes positive not before early 2002. Furthermore, the demand shock boosts output growth showing a positive contribution in early 2001 while the oil price shock contributes slightly negative in these periods. Overall, the historical decompositions show slight differences to the ones based on sign and traditional restrictions (results are displayed in Table I of Peersman, 2005). While the results appear reasonable, they still might be handled with care because of the weak validation of the non-normality assumption during the observed time period until 2007.

To avoid these sources of identification weaknesses and check robustness of the model, it might be worth to consider an extended sample until 2014Q2 including the period of the Great Recession. While this sample extension leads to non-Gaussianity of the structural shocks, we might argue that further variables are necessary to properly identify causes of economic slowdowns, in particular of the Great Recession. Furthermore, according to the replication study in Grant (2015) time varying parameter estimation might be better suited to derive at profound inferences on this extended time period. As an interesting aim for future research we leave these elaborated model modifications behind the scope of this paper.

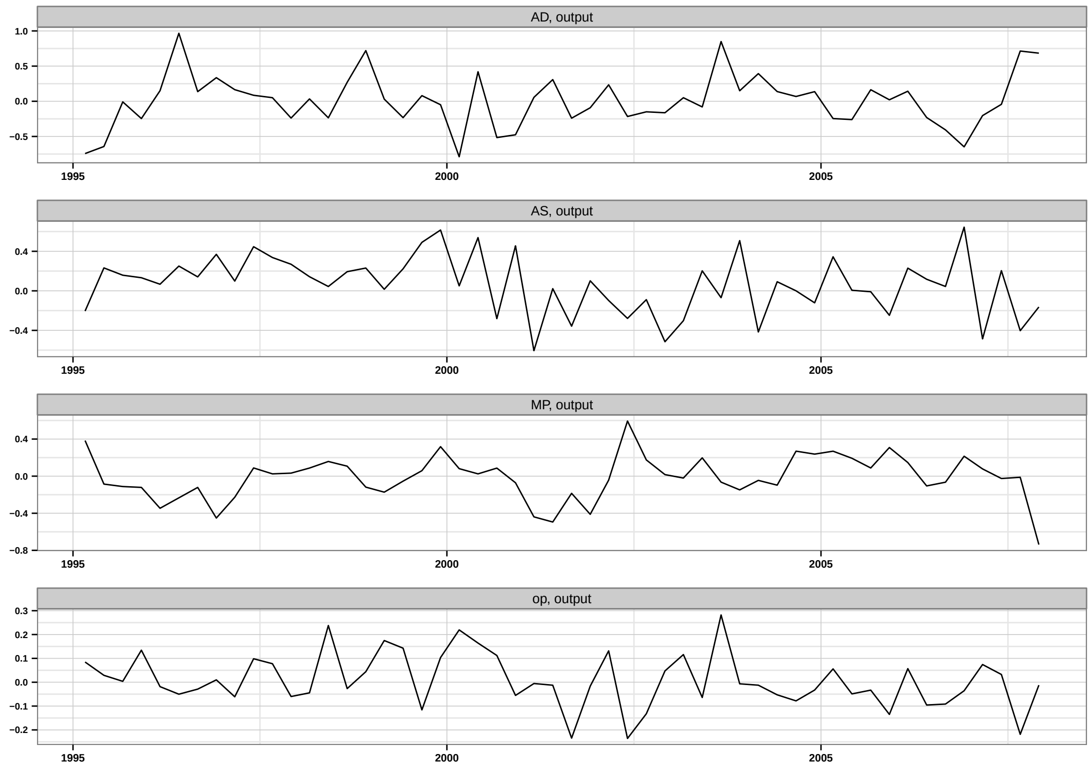


Figure 3: Historical decomposition of output growth attributed to the four shocks (oil price, aggregate supply, demand and monetary policy) based on independence based identification for sample 1995 to 2007.

4 Conclusions

Independence based identification by means of a nonparametric dependence measure allows for identification of a non-Gaussian SVAR model. We formulate identifiability in a more flexible way to overcome the limitations of this approach in the presence of multiple Gaussian structural shocks. In particular, besides identification of the whole system with at most one Gaussian component, the non-Gaussian shocks can be identified in systems which are closer to Gaussianity. Uniqueness of independence based identification of non-Gaussian structural shocks is proved theoretically. Extensions to higher dimensional systems are straightforward and might be of special interest if the analyst aims to derive economic conclusions about the response to specific shocks only (and these are non-Gaussian in their structural form).

Moreover, we retrieve these characteristics in a four dimensional macroeconomic VAR model. We revisit the study of Peersman (2005) to gain conclusive insights on macroeconomic causes of the early millennium slowdown over two different time horizons. We can uniquely identify two shocks, an oil price and a monetary policy shock, in the original sample until 2002. However, for inferences on the early millennium slowdown we advocate to consider the model ending in 2007Q4 because of non-Gaussianity of the structural shocks and a larger sample size compared to the original sample 1980Q1–2002Q2. Based on the extended sample, we obtain similar results as derived in the studies of Herwartz and Lütkepohl (2014) and Lanne and Luoto (2016). Furthermore, based on the historical decomposition of output growth into separate structural shocks we infer that a combination of shocks contributes to negative economic growth in 2001.

Appendix

Proof of Proposition 1. "$\Longleftarrow$" The proof of this implication is straightforward and, therefore, omitted.

"$\Longrightarrow$" We reformulate the $K \times K$ matrix C block wise by setting

$$C = \begin{pmatrix} C_1 & C_2 \\ C_3 & C_4 \end{pmatrix},$$

where, for instance, C_1 is a $k_1 \times k_1$ matrix. Consequently, the first k_1 Gaussian entries of ε_2 correspond to $\varepsilon_{2,1,\ldots,k_1} = \begin{pmatrix} C_1 & C_2 \end{pmatrix} \varepsilon_1$.

Suppose that one of the entries of the second block matrix C_2 would differ from zero. Following Lemma 9 of Comon (1994), the entry in ε_1 which is related to $\varepsilon_{2,1,\ldots,k_1}$ by this non zero entry in C_2 is Gaussian. This contradicts the assumption that the last $K - k_1$ components of ε_2 are non-Gaussian. Thus, $C_2 = 0_{k_1,K-k_1}$ and C_1 projects the first k_1 variables of ε_1 onto the first k_1 components of ε_2, i.e. $\varepsilon_{2,1,\ldots,k_1} = C_1\varepsilon_{1,1,\ldots,k_1}$. Assuming that the components of ε_1 are independent and its first k_1 entries are normally distributed, matrix C_1 corresponds to an orthogonal matrix Q to preserve independence of the components in $\varepsilon_{2,1,\ldots,k_1} = Q\varepsilon_{1,1,\ldots,k_1}$ (see, for instance, Hyvärinen et al., 2001).

The matrix C is assumed to be orthogonal, i.e. $CC' = I_K$. Setting $C_2 = 0_{k_1,K-k_1}$ and $C_1 = Q$

the block wise formulation of this product corresponds to

$$CC' = \begin{pmatrix} C_1 & C_2 \\ C_3 & C_4 \end{pmatrix} \begin{pmatrix} C_1' & C_3' \\ C_2' & C_4' \end{pmatrix}$$
$$= \begin{pmatrix} C_1C_1' + C_2C_2' & C_1C_3' + C_2C_4' \\ C_3C_1' + C_4C_2' & C_3C_3' + C_4C_4' \end{pmatrix} = \begin{pmatrix} QQ' & QC_3' \\ C_3Q' & C_3C_3' + C_4C_4' \end{pmatrix}.$$

Accordingly, all entries of the block matrices C_3Q' and QC_3' need to equal zero in order to obtain the identity matrix, $CC' = I_K$. As Q is orthogonal it has full rank. It follows $C_3Q' = 0_{K-k_1,k_1}$ and $QC_3' = 0_{k_1,K-k_1}$ if and only if $C_3 = 0_{K-k_1,k_1}$ with $0_{K-k_1,k_1}$ and $0_{K-k_1,k_1}$ corresponding to the $(K-k_1) \times k_1$ and $k_1 \times (K-k_1)$ zero matrices, respectively.

Hence, the product CC' can be written as

$$CC' = \begin{pmatrix} QQ' & 0 \\ 0 & C_4C_4' \end{pmatrix}.$$

Lastly, we consider the second part of ε_2 to determine the last block matrix C_4, i.e. $\varepsilon_{2,k_1+1,\dots,K} = \begin{pmatrix} 0 & C_4 \end{pmatrix} \varepsilon_1$. Matrix C_4 maps the non-Gaussian entries of ε_1 to the non-Gaussian entries of ε_2. Thus, this is an application of Comon's theorem: for independent components in ε_2, the matrix C_4 is the product of a diagonal and a permutation matrix ΛP following the derivations in Theorem 11 of Comon (1994). Finally,

$$C = \begin{pmatrix} Q & 0 \\ 0 & \Lambda P \end{pmatrix} \quad \text{and} \quad CC' = \begin{pmatrix} QQ' & 0 \\ 0 & (\Lambda P)(\Lambda P)' \end{pmatrix} = \begin{pmatrix} I_{k_1} & 0 \\ 0 & I_{K-k_1} \end{pmatrix} = I_K.$$

□

References

Blanchard, O. J., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. American Economic Review 79 (4), 655–73.

Boscolo, R., Pan, H., Roychowdhury, V. P., 2002. Beyond comon's identifiability theorem for independent component analysis. Vol. 2415 of Lecture Notes in Computer Science. Springer, pp. 1119–1124.

Chiu, C.-W. J., Mumtaz, H., Pinter, G., 2016. Var models with non-gaussian shocks. Centre for Macroeconomics (CFM) Dicussion Paper No 1609.

Comon, P., 1994. Independent component analysis, A new concept? Signal Processing 36 (3), 287–314.

Faust, J., 1998. The robustness of identified var conclusions about money. Carnegie-Rochester Conference Series on Public Policy 49 (1), 207–244.

Gavrilov, I., Pusev, R., 2015. normtest: Tests for the composite hypothesis of normality. R package version 1.1.

Gouriéroux, C., Monfort, A., Renne, J.-P., 2017. Statistical inference for independent component analysis: Application to structural var models. Journal of Econometrics 196, 111–126.

Grant, A. L., 2015. The early millennium slowdown: Replicating the peersman (2005) results. Journal of Applied Econometrics 32 (1), 224–232.

Herwartz, H., Lütkepohl, H., 2014. Structural vector autoregressions with Markov switching: combining conventional with statistical identification of shocks. J. Econometrics 183 (1), 104–116.

Herwartz, H., Plödt, M., 2016. The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. Journal of International Money and Finance 61, 30–44.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons.

Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press, forthcoming.

Lanne, M., Luoto, J., 2016. Data-driven inference on sign restrictions in bayesian structural vector autoregression. CREATES Research Paper 2016-4.

Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. J. Econometrics 196 (2), 288–304.

Lütkepohl, H., 2011. Vector autoregressive models. Eui working paper eco 2011/30, European University Institute, Florence.

Matteson, D. S., Tsay, R. S., 2017. Independent component analysis via distance covariance. Journal of the American Statistical Association 112, 623–637.

Moneta, A., Entner, D., Hoyer, P. O., Coad, A., 2013. Causal inference by independent component analysis: Theory and applications. Oxford Bulletin of Economics and Statistics 75 (5), 705–730.

Nordhausen, K., Oja, H., Tyler, D. E., Virta, J., 2016. ICtest: Estimating and Testing the Number of Interesting Components in Linear Dimension Reduction. R package version 0.2.

Peersman, G., 2005. What caused the early millennium slowdown? Evidence based on vector autoregressions. J. Appl. Econometrics 20 (2), 185–207.

Risk, B. B., James, N. A., Matteson, D. S., 2015. steadyICA: ICA and Tests of Independence via Multivariate Distance Covariance. R package version 1.0.

Sims, C. A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.

Székely, G. J., Rizzo, M. L., Bakirov, N. K., 2007. Measuring and testing dependence by correlation of distances. Ann. Statist. 35 (6), 2769–2794.

Uhlig, H., 2005. What are the effects of monetary policy on output? results from an agnostic identification procedure. Journal of Monetary Economics 52 (2), 381–419.