Subvector inference in GARCH-X type models

Philipp Ketz*

January 11, 2022

Preliminary and incomplete. Please do not circulate.

In this paper, we derive asymptotic distribution theory for a general class of models where the identification strength of one parameter is determined by another parameter and where the latter is allowed to be at the boundary of the parameter space—extending the results in Andrews and Cheng (2012). This allows us to derive the asymptotic distribution, under different identification strengths, of the two test statistics that are used in the two-step (testing) procedure proposed in Pedersen and Rahbek (2019). The latter aims at testing the null hypothesis that a GARCH-X type model, with an exogenous covariate (X), reduces to a standard GARCH model, while allowing the "GARCH parameter" to be unidentified. We find that using the second step test statistic together with plug-in least favorable configuration (PI-LF) critical values offers (asymptotic) power gains over a wide range of alternatives (for realistic choices of the data generating process) compared to the two-step procedure. Furthermore, we find that the two-step procedure fails to control asymptotic size. Together, our findings provide arguments against the use of the two-step procedure in practice; other tests, such as the aforementioned test using PI-LF critical values, ought to be preferred.

Keywords: Boundary, lack of identification, testing.

*Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France. Email: philipp.ketz@psemail.eu.

1 Introduction

In GARCH-X type models, the variance of a generalized autoregressive conditional heteroskedasticity (GARCH) type model is augmented by a set of "exogenous" regressors (X). Naturally, the question arises if the more general GARCH-X type model reduces to the simpler GARCH type model. Statistically speaking, the problem reduces to testing whether the "coefficients" on the exogenous regressors are significantly different from zero. The testing problem is non-standard due to the presence of two nuisance parameters that could possibly be at the boundary of the parameter space. In addition, under the null hypothesis, one of the nuisance parameters, the "GARCH parameter," is not identified when the other, the "ARCH parameter," is at the boundary. In order to address this possible lack of identification, Pedersen and Rahbek (2019) (PR hereinafter) suggest a two-step testing procedure, where rejection in the first step is taken as "evidence" that the model is strongly identified. In the second step, the authors then rely on an additional assumption, namely that a specific entry of the inverse information is zero, to obtain an asymptotic null distribution of their proposed ("second step") test statistic that is nuisance parameter free. There are two (potential) problems with this two-step procedure. First, as is well known, two-step procedures may suffer from asymptotic size distortions, i.e., the asymptotic size may exceed the nominal level, if the possible error in the first step is not appropriately taken into account (see e.g., Leeb and Pötscher, 2005, 2008). In addition, the aforementioned additional assumption that the authors make in the second step may not be satisfied (under reasonable assumptions about the data generating process/true parameter space), which may lead to or aggravate ("existing") asymptotic size distortions. Second, the two-step procedure may, irrespective of possible asymptotic size distortions, have poor (asymptotic) power over large parts of the parameter space, as suggested by simulations in PR. In sum, the two-step procedure may suffer from large type I and type II errors in different parts of the parameter space.

The motivation for the two-step procedure comes from the fact that the asymptotic null distribution of the "second step" test statistic (and the underlying estimator) under semi-strong and weak identification (of the "GARCH parameter"), using the terminology in Andrews and Cheng (2012) (AC hereinafter), is unknown in the literature to date. In this paper, we fill this gap in the literature by extending the results in AC to allow for the parameter that determines the identification strength to be at the boundary of the parameter space such that they cover the GARCH-X type models considered in PR. This, in turn, allows us to characterize the asymptotic size (AsySz) of the two-step procedure proposed in PR in terms of a finite number of parameters—we are currently in the process of numerically evaluating the AsySz in order to establish whether the procedure controls asymptotic size or not. Furthermore, the new results open the possibility to implement (one-step) testing procedures that, by construction, control asymptotic size and that, in addition, have good power properties. For example, the "second step" test statistic used in PR together with plug-in least favorable configuration (PI-LF) critical values is found to have greater (asymptotic) power than the two-step procedure over a large part of the parameter space (for realistic choices of the data generating process), illustrating the potential for improvement.

The plan of this paper is as follows. In Section 2, we present the new asymptotic distribution theory and introduce the GARCH-X model, which serves as running example. Section 3 introduces the test statistics used in the two-step procedure proposed by PR. In Section 5, we graphically illustrate that the finite-sample distribution of the test statistics and the underlying estimators is well approximated by the asymptotic distribution theory. In addition, this section provides an asymptotic power comparison of the two-step procedure proposed by PR and the test that uses their "second step" test statistic together with PI-LF critical values. Section 6 investigates the asymptotic size properties of the two-step procedure and Section 7 concludes.

2 Asymptotic theory

This section closely follows AC. Let $\theta = (\psi', \pi')' = (\beta', \zeta', \pi')'$ denote the finite-dimensional parameter of interest, where β governs the identification strength of π and $\psi = (\beta', \zeta')'$ is always identified. We consider the estimator $\hat{\theta}_n$ that satisfies $\hat{\theta}_n \in \Theta$ and

$$Q_n(\hat{\theta}_n) = \inf_{\theta \in \Theta} Q_n(\theta) + o(1),$$

where Θ and $Q_n(\theta)$ denote the "optimization" parameter space and the objective function, respectively. The dependence of $Q_n(\theta)$ on the data $\{W_t : t \leq n\}$ is suppressed for notational convenience. The true data generating process is indexed by $\gamma^* = (\theta^*, \phi^*)$ and the corresponding "true parameter space" is denoted Γ. Here, θ^* denotes the true value of θ and ϕ^*, which denotes the true value of ϕ, indexes the part of the distribution of the data not determined by θ^*, such as the distribution of the conditioning variables in conditional maximum likelihood (CML). We assume that π is unidentified when $\beta = 0$.

Assumption A. If $\beta = 0$, $Q_n(\theta)$ does not depend on π $\forall \theta = (\beta, \zeta, \pi) = (0, \zeta, \pi) \in \Theta, \forall n \geq 1$, for any true parameter $\gamma^* \in \Gamma$.

Running example - GARCH-X(1,1): We consider a simple version of the GARCH-X(1,1) model, with a single exogenous variable (as in PR). The model is given by $y_t = h_t(\theta^*)^{1/2} z_t$, where

$$h_t(\theta) = h_t(\psi, \pi) = h_t(\beta, \zeta, \pi) = \zeta(1-\pi) + \beta_1 y_{t-1}^2 + \pi h_{t-1}(\psi, \pi) + \beta_2 x_{t-1}^2$$

and where ϕ^* denotes the distribution of $\{z_t, x_t\}$. Here, y_t and x_t are observed and z_t is unobserved. Note that $\beta = (\beta_1, \beta_2)' \in \mathbb{R}^2$, while $\zeta, \pi \in \mathbb{R}$. The objective function is given by (-$\frac{1}{n}$ times) the Gaussian-based conditional quasi log-likelihood function, i.e., $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^n l_t(\theta)$, where

$$l_t(\theta) = \frac{1}{2}\log(2\tilde{\pi}) + \frac{1}{2}\log(h_t(\theta)) + \frac{y_t^2}{2h_t(\theta)}$$

and where $\tilde{\pi} = 3.14...$, with initial condition $h_0(\theta) = \zeta$. The optimization parameter space, Θ, is given by $\Theta = \Psi \times \Pi$, where

$$\Psi = \{\psi : 0 \leq \beta_1 \leq \overline{\beta}_1, 0 \leq \beta_2 \leq \overline{\beta}_2, \underline{\zeta} \leq \zeta \leq \overline{\zeta}\} \text{ and } \Pi = \{\pi : 0 \leq \pi \leq \overline{\pi}\}$$

for some $0 < \overline{\beta}_1, \overline{\beta}_2 < \infty$, $0 < \underline{\zeta} < \overline{\zeta} < \infty$, and $0 < \overline{\pi} < 1$. Note that, given $h_0(\theta) = \zeta$, we have

$$h_t(\theta) = \zeta + \beta_1 \sum_{i=0}^{t-1} \pi^i y_{t-i-1}^2 + \beta_2 \sum_{i=0}^{t-1} \pi^i x_{t-i-1}^2$$

and $h_t(0, \zeta, \pi) = \zeta \; \forall \theta = (\beta, \zeta, \pi) = (0, \zeta, \pi) \in \Theta, \forall n \geq 1$, such that Assumption A is satisfied.

The true distribution of the data $\{W_t : t \leq n\}$ is denoted F_γ where $\gamma \in \Gamma$. P_γ and E_γ denote probability and expectation under γ, respectively. The true parameter space, Γ, is assumed to be compact and of the following form

$$\Gamma = \{\gamma = (\theta, \phi) : \theta \in \Theta^*, \phi \in \Phi(\theta)\},$$

where Θ^* is compact and $\Phi(\theta) \subset \Phi \; \forall \theta \in \Theta^*$ for some compact metric space Φ with a metric that induces weak convergence of the distribution of $(W_{n,t}, W_{n,t+m})$ for all $i, m \geq 1$, i.e., the metric is such that if $\gamma^1 \to \gamma^2$, then $(W_{n,i}, W_{n,i+m})$ under γ^1 converges in distribution to $(W_{n,i}, W_{n,i+m})$ under γ^2 for all $i, m \geq 1$.

Running example continued: The true parameter space Θ^* is given by $\Psi^* \times \Pi^*$, where

$$\Psi^* = \{\psi : 0 \leq \beta_1 \leq \overline{\beta}_1^*, 0 \leq \beta_2 \leq \overline{\beta}_2^*, \underline{\zeta}^* \leq \zeta \leq \overline{\zeta}^*\} \text{ and } \Pi^* = \{\pi : 0 \leq \pi \leq \overline{\pi}^*\}$$

for some $0 < \overline{\beta}_1^* < \overline{\beta}_1$, $0 < \overline{\beta}_2^* < \overline{\beta}_2$, $\underline{\zeta} < \underline{\zeta}^* < \overline{\zeta}^* < \overline{\zeta}$, and $0 < \overline{\pi}^* < \overline{\pi}$. Given this definition of the parameter space, boundary effects are permitted for β_1, β_2, and π, but only "below" at zero. [To do: Define $\Phi(\theta)$ and Φ.]

In this paper, we are interested in the behavior of tests under varying identification strength. Say we are interested in testing $H_0 : r(\theta) = v$ for some function $r(\theta)$ of θ. Let $\mathcal{T}_n(v)$ denote a generic test statistic for testing $H_0 : r(\theta) = v$ and let $\mathrm{cv}_{n,1-\alpha}(v)$ denote the corresponding nominal level α critical value, which may depend on n and v. We approximate the (finite-sample) size of the resulting test, $\sup_{\gamma\in\Gamma:r(\theta)=v} P_\gamma(\mathcal{T}_n(v) > \mathrm{cv}_{n,1-\alpha}(v))$, by its asymptotic size (AsySz), i.e.,

$$\mathrm{AsySz} = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma:r(\theta)=v} P_\gamma(\mathcal{T}_n(v) > \mathrm{cv}_{n,1-\alpha}(v)).$$

In order to characterize AsySz and eventually determine whether a given test controls asymptotic size, i.e., AsySz $\leq \alpha$, we rely on Lemma 2.1 in AC (adapted to tests). Therefore, we consider the same set of drifting sequences of true parameters $\gamma_n = (\theta_n, \phi_n)$ as AC, see their equation (2.7).

2.1 Corresponds to Section B in AC

[To do: "Paste" Assumptions B1–B3 from AC, suitably adapting Assumption B(i) to allow for "some" boundary effects.]

2.2 Corresponds to Section C in AC

[To do: Complete and re-formulate this section.]

Define the concentrated extremum estimator $\hat{\psi}_n(\pi)(\in \Psi(\pi))$ of ψ for given $\pi \in \Pi$ by

$$Q_n(\hat{\psi}_n(\pi), \pi) = \inf_{\psi\in\Psi(\pi)} Q_n(\psi, \pi) + o(n^{-1}).$$

Below, we will use the following notation: $Q_n^c(\pi) = Q_n(\hat{\psi}_n(\pi), \pi)$. We rely on the same definitions of sets of drifting sequences as those used in AC. Let

$$a_n(\gamma_n) = \begin{cases} n^{1/2} & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } \|b\| < \infty \\ \|\beta_n\|^{-1} & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } \|b\| = \infty \end{cases}.$$

We invoke Assumptions C1–C5. In particular (given Assumptions C1 and C4), we have the following quadratic expansion in ψ around $\psi_{0,n} = (0, \zeta_n)$ for given π, under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$:

$$Q_n(\psi,\pi) = Q_n(\psi_{0,n},\pi) + D_\psi Q_n(\psi_{0,n},\pi)'(\psi - \psi_{0,n}) \\ + \frac{1}{2}(\psi - \psi_{0,n})' H(\pi;\gamma_0)(\psi - \psi_{0,n}) + R_n^*(\psi,\pi).$$

Running example - continued: Assumption C1 holds with

$$D_\psi Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \rho_{\psi,t}(\theta),$$

where

$$\rho_{\psi,t}(\theta) \equiv \frac{\partial}{\partial\psi}\rho_t(\theta) = \frac{1}{2h_t^\infty(\theta)}\left(1 - \frac{y_t^2}{h_t^\infty(\theta)}\right)\frac{\partial h_t^\infty(\theta)}{\partial\psi}$$

and where

$$\frac{\partial h_t^\infty(\theta)}{\partial\psi} = \left(\sum_{i=0}^{\infty}\pi^i y_{t-i-1}^2, \sum_{i=0}^{\infty}\pi^i x_{t-i-1}^2, 1\right)'.$$

It will be convenient to introduce $H(\pi;\gamma_0)$ later on. Assumptions C2 and C3 hold with

$$G_n(\pi) = \sqrt{n}\frac{1}{n}\sum_{t=1}^{n}(\rho_{\psi,t}(\psi_{0,n},\pi) - E_{\gamma_n}\rho_{\psi,t}(\psi_{0,n},\pi)),$$

where $G_n(\cdot) \xrightarrow{d} G(\cdot;\gamma_0)$ and where $G(\cdot;\gamma_0)$ denotes a mean zero Gaussian process with covariance Kernel given by

$$\Omega(\pi_1,\pi_2;\gamma_0) = E_{\gamma_0}\rho_{\psi,t}(\psi_0,\pi_1)\rho_{\psi,t}(\psi_0,\pi_2)',$$

(under appropriate assumptions about the dependence of z_t (or maybe y_t!?) and $\mathcal{F}_{t-1}$). Note that, here, γ_0 is such that $\beta_0 = 0$. Furthermore, note that, if the true parameter is γ^*, then

$$\frac{y_t^2}{h_t^\infty(\theta^*)} = z_t^2.$$

Therefore, if γ^* is such that $\beta^* = 0$, then $\frac{y_t^2}{\zeta^*} = z_t^2$. Letting $c \equiv \frac{E_{\gamma_0}(1-z_t^2)^2}{2} = \frac{E_{\gamma_0}z_t^4 - 1}{2}$ (since $E_{\gamma_0}[z_t^2|\mathcal{F}_{t-1}] = E_{\gamma_0}[z_t^2] = 1$) following Andrews (2001) [See also equation (4.4) in Andrews (2001); Note that we impose Assumption 2.2 in Pedersen and Rahbek (2019)], we have the following simplification:

$$\Omega(\pi_1,\pi_2;\gamma_0) = \frac{c}{2\zeta_0^2}E_{\gamma_0}\frac{\partial h_t^\infty(\psi_0,\pi_1)}{\partial\psi}\frac{\partial h_t^\infty(\psi_0,\pi_2)}{\partial\psi'}$$

We have several more simplifications. First,

$$\begin{aligned}
\frac{1}{\zeta_0^2} E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i y_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j y_{t-j-1}^2 &= E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i z_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j z_{t-j-1}^2 \\
&= \sum_{i=0}^{\infty} \pi_1^i \pi_2^i E_{\gamma_0} z_{t-i-1}^4 + \sum_{i=0}^{\infty} \sum_{j \neq i} \pi_1^i \pi_2^j E_{\gamma_0} z_{t-i-1}^2 z_{t-j-1}^2 \\
&= \sum_{i=0}^{\infty} \pi_1^i \pi_2^i E_{\gamma_0} z_{t-i-1}^4 - \sum_{i=0}^{\infty} \pi_1^i \pi_2^i + \sum_{i=0}^{\infty} \pi_1^i \sum_{j=0}^{\infty} \pi_2^j \\
&= \frac{E_{\gamma_0} z_t^4 - 1}{1 - \pi_1 \pi_2} + \frac{1}{(1-\pi_1)(1-\pi_2)} \\
&= \frac{2c}{1 - \pi_1 \pi_2} + \frac{1}{(1-\pi_1)(1-\pi_2)},
\end{aligned}$$

where the second to last equality uses $E_{\gamma_0} z_{t-i-1}^2 z_{t-j-1}^2 = 1$ for $i \neq j$. Second,

$$\frac{1}{\zeta_0} E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i y_{t-i-1}^2 = E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i z_{t-i-1}^2 = \sum_{i=0}^{\infty} \pi_1^i E_{\gamma_0} z_{t-i-1}^2 = \frac{1}{1-\pi_1}.$$

Third,

$$\frac{1}{\zeta_0} E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i y_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2 = E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i z_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2.$$

[Note that, if we are willing to assume independence, this equals $\frac{1}{1-\pi_1} E_{\gamma_0} \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2$.] In sum, we have

$$\Omega(\pi_1, \pi_2; \gamma_0) = \frac{c}{2} \begin{bmatrix} \frac{2c}{1-\pi_1\pi_2} + \frac{1}{(1-\pi_1)(1-\pi_2)} & \frac{1}{\zeta_0} E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i z_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2 & \frac{1}{\zeta_0} \frac{1}{1-\pi_1} \\ \frac{1}{\zeta_0} E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i x_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j z_{t-j-1}^2 & \frac{1}{\zeta_0^2} E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i x_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2 & \frac{1}{\zeta_0^2} E_{\gamma_0} \sum_{j=0}^{\infty} \pi_1^j x_{t-j-1}^2 \\ \frac{1}{\zeta_0} \frac{1}{1-\pi_2} & \frac{1}{\zeta_0^2} E_{\gamma_0} \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2 & \frac{1}{\zeta_0^2} \end{bmatrix}.$$

By definition, we have

$$H(\pi; \gamma_0) = E_{\gamma_0} \frac{\partial}{\partial \psi'} \rho_{\psi,t}(\psi_0, \pi).$$

Also,

$$\frac{\partial}{\partial \psi'} \rho_{\psi,t}(\theta) = \frac{\partial}{\partial \psi'} \left[\frac{1}{2h_t^{\infty}(\theta)} \left(1 - \frac{y_t^2}{h_t^{\infty}(\theta)}\right) \frac{\partial h_t^{\infty}(\theta)}{\partial \psi} \right] = \frac{\partial h_t^{\infty}(\theta)}{\partial \psi} \frac{\partial}{\partial \psi'} \left[\frac{1}{2h_t^{\infty}(\theta)} \left(1 - \frac{y_t^2}{h_t^{\infty}(\theta)}\right) \right],$$

since $\frac{\partial h_t^{\infty}(\theta)}{\partial \psi}$ is not a function of ψ. We have

$$\frac{\partial}{\partial \psi'} \left[\frac{1}{2h_t^{\infty}(\theta)} \left(1 - \frac{y_t^2}{h_t^{\infty}(\theta)}\right) \right] = \left[-\frac{1}{2(h_t^{\infty}(\theta))^2} \left(1 - \frac{y_t^2}{h_t^{\infty}(\theta)}\right) + \frac{1}{2h_t^{\infty}(\theta)} \frac{y_t^2}{(h_t^{\infty}(\theta))^2} \right] \frac{\partial h_t^{\infty}(\theta)}{\partial \psi'}.$$

Now, using $E_{\gamma_0} z_t^2 = 1$, we have

$$E_{\gamma_0} - \frac{1}{2\zeta_0^2}\left(1 - z_t^2\right)\frac{\partial h_t^\infty(\psi_0,\pi)}{\partial \psi}\frac{\partial h_t^\infty(\psi_0,\pi)}{\partial \psi'} = 0.$$

Therefore, we have

$$H(\pi;\gamma_0) = \frac{1}{2\zeta_0^2} E_{\gamma_0}\frac{\partial h_t^\infty(\psi_0,\pi)}{\partial \psi}\frac{\partial h_t^\infty(\psi_0,\pi)}{\partial \psi'} = \Omega(\pi,\pi;\gamma_0)/c.$$

Next, we derive $K(\pi;\gamma_0) \equiv K(\psi_0,\pi;\gamma_0)$ defined in Assumption C5. By definition, we have

$$K_n(\theta;\gamma^*) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial \beta^{*\prime}} E_{\gamma^*}\rho_{\psi,t}(\theta).$$

We have

$$\begin{aligned}
E_{\gamma^*}\rho_{\psi,t}(\theta) &= E_{\gamma^*}\frac{1}{2h_t^\infty(\theta)}\left(1 - \frac{y_t^2}{h_t^\infty(\theta)}\right)\frac{\partial h_t^\infty(\theta)}{\partial \psi} = E_{\gamma^*}\frac{1}{2h_t^\infty(\theta)}\left(1 - \frac{h_t^\infty(\theta^*)}{h_t^\infty(\theta)}\right)\frac{\partial h_t^\infty(\theta)}{\partial \psi} \\
&= E_{\gamma^*}\frac{1}{2h_t^\infty(\theta)}\frac{\partial h_t^\infty(\theta)}{\partial \psi} - E_{\gamma^*}\frac{1}{2h_t^\infty(\theta)}\frac{h_t^\infty(\theta^*)}{h_t^\infty(\theta)}\frac{\partial h_t^\infty(\theta)}{\partial \psi} \\
&= E_{\gamma^*}\frac{1}{2h_t^\infty(\theta)}\frac{\partial h_t^\infty(\theta)}{\partial \psi} - E_{\gamma^*}\frac{1}{2h_t^\infty(\theta)}\frac{\zeta^* + \beta_1^*\sum_{i=0}^\infty \pi^{*i} y_{t-i-1}^2 + \beta_2^*\sum_{i=0}^\infty \pi^{*i} x_{t-i-1}^2}{h_t^\infty(\theta)}\frac{\partial h_t^\infty(\theta)}{\partial \psi}.
\end{aligned}$$

Then, we have

$$\begin{aligned}
K(\pi;\gamma_0) &= -E_{\gamma_0}\frac{1}{2\zeta_0^2}\left.\frac{\partial h_t^\infty(\theta)}{\partial \psi}\right|_{\theta=(\psi_0,\pi)}\left[\sum_{j=0}^\infty \pi_0^j y_{t-j-1}^2 ; \sum_{j=0}^\infty \pi_0^j x_{t-j-1}^2\right] \\
&= -\frac{1}{2}\begin{bmatrix}
\frac{2c}{1-\pi\pi_0} + \frac{1}{(1-\pi)(1-\pi_0)} & \frac{1}{\zeta_0}E_{\gamma_0}\sum_{i=0}^\infty \pi^i z_{t-i-1}^2 \sum_{j=0}^\infty \pi_0^j x_{t-j-1}^2 \\
\frac{1}{\zeta_0}E_{\gamma_0}\sum_{i=0}^\infty \pi^i x_{t-i-1}^2 \sum_{j=0}^\infty \pi_0^j z_{t-j-1}^2 & \frac{1}{\zeta_0^2}E_{\gamma_0}\sum_{i=0}^\infty \pi^i x_{t-i-1}^2 \sum_{j=0}^\infty \pi_0^j x_{t-j-1}^2 \\
\frac{1}{\zeta_0}\frac{1}{1-\pi_0} & \frac{1}{\zeta_0^2}E_{\gamma_0}\sum_{j=0}^\infty \pi_0^j x_{t-j-1}^2
\end{bmatrix}
\end{aligned}$$

Define

$$Z_n(\pi) = -a_n(\gamma_n)H^{-1}(\pi;\gamma_0)D_\psi Q_n(\psi_{0,n},\pi).$$

Let

$$q_n(\lambda,\pi;\gamma_0) = (\lambda - Z_n(\pi))' H(\pi;\gamma_0)(\lambda - Z_n(\pi)).$$

Then, we have

$$Q_n(\psi,\pi) = Q_n(\psi_{0,n},\pi) - \frac{1}{2a_n^2(\gamma_n)} Z_n(\pi)' H(\pi;\gamma_0) Z_n(\pi) + \frac{1}{2a_n^2(\gamma_n)} q_n(a_n(\gamma_n)(\psi - \psi_{0,n}),\pi;\gamma_0) + R_n^*(\psi,\pi).$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, we have

$$Z_n(\pi) \xrightarrow{d} \begin{cases} Z(\pi;\gamma_0,b) \equiv -H^{-1}(\pi;\gamma_0)\{G(\pi;\gamma_0) + K(\pi;\gamma_0)b\} & \text{if } \|b\| < \infty \\ -H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0)\omega_0 & \text{if } \|b\| = \infty \text{ and } \beta_n/\|\beta_n\| \to \omega_0 \end{cases},$$

cf. equation (3.7) in Andrews and Cheng (2012). Define

$$q(\lambda,\pi;\gamma_0,b) = (\lambda - Z(\pi;\gamma_0,b))' H(\pi;\gamma_0)(\lambda - Z(\pi;\gamma_0,b))$$

and

$$\hat{\lambda}(\pi;\gamma_0,b) = \arg\min_{\lambda \in \Lambda} q(\lambda,\pi;\gamma_0,b),$$

where *in our running example*

$$\Lambda = [0,\infty]^2 \times [-\infty,\infty].$$

Then the equivalent for the first part of Theorem 3.1(a) for *fixed* π is given by

$$\sqrt{n}(\hat{\psi}_n(\pi) - \psi_{0,n}) \xrightarrow{d} \hat{\lambda}(\pi;\gamma_0,b).$$

The "rest" of part Theorem 3.1(a) also holds, with $\pi^*(\gamma_0,b)$ suitably adapted, which is necessary because Assumption C6 does not (/cannot) hold in the presence of boundary constraints. This is an immediate consequence of Assumption A. To see this note that (*in our running example*) $\hat{\lambda}(\pi;\gamma_0,b)_{1:2} = (0,0)'$ with positive probability (QUESTION: is it always true that this holds for all π??? No, not at all!!! But I don't think that this is a problem per se.) under $\{\gamma_n\} \in \Gamma(\gamma_0,0,b)$ with $\|b\| < \infty$ such that

$$Q_n^c(\pi) = Q_{0,n} + o_p(1),$$

where $Q_{0,n} = Q_n(\psi_{0,n},\pi)$, which does not depend on π. To "fix" this, we have to modify Assumption C6 so that it holds whenever $\hat{\lambda}(\pi;\gamma_0,b)_{1:2} > (0,0)'$ and $\pi^*(\gamma_0,b)$ can be defined arbitrarily whenever $\hat{\lambda}(\pi;\gamma_0,b)_{1:2} = (0,0)'$ for all π. Similarly, the equivalent of Theorem

3.1(b) is given by

$$2n(Q_n(\hat{\theta}_n) - Q_{0,n}) \xrightarrow{d} \inf_{\pi \in \Pi} -\hat{\lambda}(\pi;\gamma_0,b)' H(\pi;\gamma_0)\hat{\lambda}(\pi;\gamma_0,b).$$

Note that the *fixed* π version in Lemma 3.2(a) holds as well. Lemma 3.2(b) also holds (without modification). To see this, note that under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| = \infty$ and $\beta_n/\|\beta_n\| \to \omega_0$, the equivalent of the above $q(\cdot)$ function is given by

$$(\lambda - \{-H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0)\omega_0\})' H(\pi;\gamma_0)(\lambda - \{-H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0)\omega_0\}),$$

which we minimize over

$$\frac{1}{\|\beta_n\|}(\Psi - \psi_{0,n}) \to \Lambda$$

(at least *in our running example*). Furthermore, *in our running example*, $-H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0)\omega_0$ is strictly positive (and finite), such that the equivalent of the $q(\cdot)$ function equals 0.

[To do: Verification of Assumptions C7 and C8 and Lemmas 3.3 and 3.4.]

2.3 Corresponds to Section D in AC

[To do: "Reproduce" suitably adapted versions of Assumptions D1–D3.]

Running example - continued: We want to derive $J(\gamma_0)$ and $V(\gamma_0)$ that pop up in Assumptions D2 and D3. Let's first consider D3

$$B^{-1}(\beta_n)DQ_n(\theta_n) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{2h_t^{\infty}(\theta_n)}\left(1 - \frac{y_t^2}{h_t^{\infty}(\theta_n)}\right)\varphi(\theta_n),$$

where

$$\varphi(\theta_n) = B^{-1}(\beta_n)\frac{\partial h_t^{\infty}(\theta_n)}{\partial \theta} = \left(\sum_{i=0}^{\infty} \pi_n^i y_{t-i-1}^2, \sum_{i=0}^{\infty} \pi_n^i x_{t-i-1}^2, 1, \sum_{i=1}^{\infty} i\pi_n^{i-1}\left(\frac{\beta_{n,1}}{\|\beta_n\|} y_{t-i-1}^2 + \frac{\beta_{n,2}}{\|\beta_n\|} x_{t-i-1}^2\right)\right)'.$$

We have

$$V(\gamma_0) = \frac{c}{2} E_{\gamma_0} \frac{\varphi(\theta_0)}{h_t^{\infty}(\theta_0)} \frac{\varphi'(\theta_0)}{h_t^{\infty}(\theta_0)}.$$

Similarly, using similar arguments as in Lemma A.6 in Pedersen and Rahbek (2019), we can show that

$$J(\gamma_0) = \frac{1}{2} E_{\gamma_0} \frac{\varphi(\theta_0)}{h_t^{\infty}(\theta_0)} \frac{\varphi'(\theta_0)}{h_t^{\infty}(\theta_0)}.$$

3 Test statistics used in two-step procedure

Pedersen and Rahbek (2019) consider two test statistics. Assuming that $\hat{\theta}_n = (\hat{\psi}_n(\hat{\pi}_n), \hat{\pi}_n)$, where $\hat{\pi}_n \in \Pi$ is defined as

$$Q_n^c(\hat{\pi}_n) = \inf_{\pi \in \Pi} Q_n^c(\pi) + o(n^{-1}),$$

the LR^* statistic for testing $H_0^* : \beta_1 = \beta_2 = 0$ can be written as

$$LR^* = 2n(\min_{\theta \in \Theta_0^*} Q_n(\theta) - Q_n(\hat{\theta}_n)),$$

where $\Theta_0^* = \{\theta \in \Theta : \beta_1 = \beta_2 = 0\}$.

The LR statistic for testing $H_0^* : \beta_2 = 0$ can be written as

$$LR = 2n(\min_{\theta \in \Theta_0} Q_n(\theta) - Q_n(\hat{\theta}_n)),$$

where $\Theta_0 = \{\theta \in \Theta : \beta_2 = 0\}$. Let $S_\beta = [I_2 \ 0_2]$ denote the selection matrix that selects the entries pertaining to β and let $\hat{\lambda}_\beta(\pi; \gamma_0, b) = S_\beta \hat{\lambda}(\pi; \gamma_0, b)$. Then, the asymptotic distribution of LR^* under $\beta_n = b/\sqrt{n}$ is given by

$$\sup_{\pi \in \Pi} \hat{\lambda}_\beta(\pi; \gamma_0, b)'(S_\beta H^{-1}(\pi; \gamma_0) S_\beta')^{-1} \hat{\lambda}_\beta(\pi; \gamma_0, b)(-0_2'(S_\beta H^{-1}(\pi; \gamma_0) S_\beta')^{-1} 0_2).$$

Note that under $\beta_n = 0_2$ $(\forall n)$, we have $b = 0$.

The asymptotic distribution of LR under $\beta_n = b/\sqrt{n}$ is given by

$$\sup_{\pi \in \Pi} \hat{\lambda}_\beta(\pi; \gamma_0, b)'(S_\beta H^{-1}(\pi; \gamma_0) S_\beta')^{-1} \hat{\lambda}_\beta(\pi; \gamma_0, b) - \sup_{\pi \in \Pi} \hat{\lambda}_\beta^r(\pi; \gamma_0, b)'(S_\beta H^{-1}(\pi; \gamma_0) S_\beta')^{-1} \hat{\lambda}_\beta^r(\pi; \gamma_0, b),$$

where $\hat{\lambda}_\beta^r(\pi; \gamma_0, b) = S_\beta \hat{\lambda}^r(\pi; \gamma_0, b)$,

$$\hat{\lambda}_\beta^r(\pi; \gamma_0, b) = \arg\min_{\lambda \in \Lambda^r} q(\lambda, \pi; \gamma_0, b),$$

and

$$\Lambda^r = [0, \infty] \times \{0\} \times [-\infty, \infty].$$

If we assume that

$$E_{\gamma_0} \sum_{i=0}^{\infty} \pi_1^i z_{t-i-1}^2 \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2 = \frac{1}{1 - \pi_1} E_{\gamma_0} \sum_{j=0}^{\infty} \pi_2^j x_{t-j-1}^2 \tag{1}$$

(a sufficient condition is that z_t ($t = ..., -2, -1, 0, 1, 2, ...$) is independent of x_t), then $S_\beta H^{-1}(\pi; \gamma_0) S'_\beta$ is diagonal. In that case, the two asymptotic distributions simplify: Let $S_{\beta_1} = [1\ 0\ 0]$ and $S_{\beta_2} = [0\ 1\ 0]$ denote the selection matrices that select the entries pertaining to β_1 and β_2, respectively. Then, $\hat{\lambda}_{\beta_1}(\pi; \gamma_0, b) = S_{\beta_1}\hat{\lambda}(\pi; \gamma_0, b)$ and $\hat{\lambda}_{\beta_2}(\pi; \gamma_0, b) = S_{\beta_2}\hat{\lambda}(\pi; \gamma_0, b)$, where, for $i \in \{1, 2\}$,

$$\hat{\lambda}_{\beta_i}(\pi; \gamma_0, b) \sim \max(Z_{\beta_i}(\pi; \gamma_0, b), 0),$$

with $Z_{\beta_i}(\pi; \gamma_0, b) = S_{\beta_i} Z(\pi; \gamma_0, b)$. Let $\Delta = (S_\beta H^{-1}(\pi; \gamma_0) S'_\beta)^{-1}$. Then, the asymptotic distribution of LR is, for example, given by

$$\sup_{\pi \in \Pi} \left[\hat{\lambda}_{\beta_1}(\pi; \gamma_0, b)^2 \Delta_{11} + \hat{\lambda}_{\beta_2}(\pi; \gamma_0, b)^2 \Delta_{22} \right] - \sup_{\pi \in \Pi} \hat{\lambda}_{\beta_1}(\pi; \gamma_0, b)^2 \Delta_{11}. \tag{2}$$

3.1 Plug-in least favorable configuration

In the case where equation (2) applies, applying plug-in least favorable configuration critical values (AC) (for the LR statistic) is not too computationally expensive. In addition, it should be possible to exploit certain monotonicities. [To do: Insert the computation of the plug-in least favorable configuration critical values, for when equation (2) applies and for when it doesn't.]

4 Approximation

The *dgp* is generated as follows: $n = 500$. Burn-in phase of 100 observations. We are generating data under the null, i.e., $\beta_2 = 0$. Furthermore, we choose $\zeta = 1$ (which has no impact it seems) and $\pi = 0.2$. We vary β_1 in the set $\{0, 0.1, 0.2\}$ which correspond to the localization parameter $b_1 = \sqrt{500} \cdot \beta_1 \in \{0, 2.23, 4.47\}$. x_t is drawn from an AR(1), i.e.,

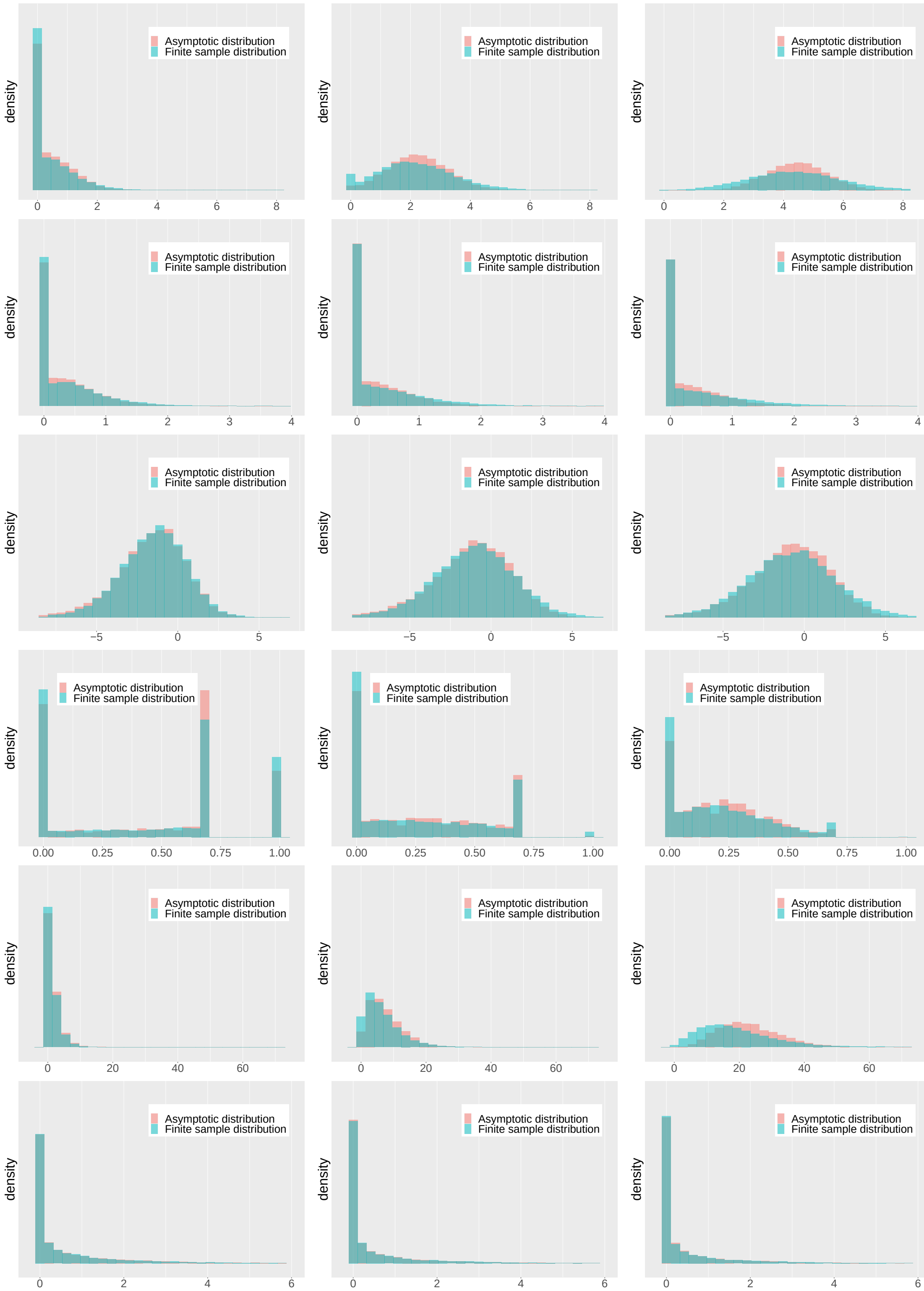
$$x_t = \rho_x x_{t-1} + \epsilon_t,$$

where

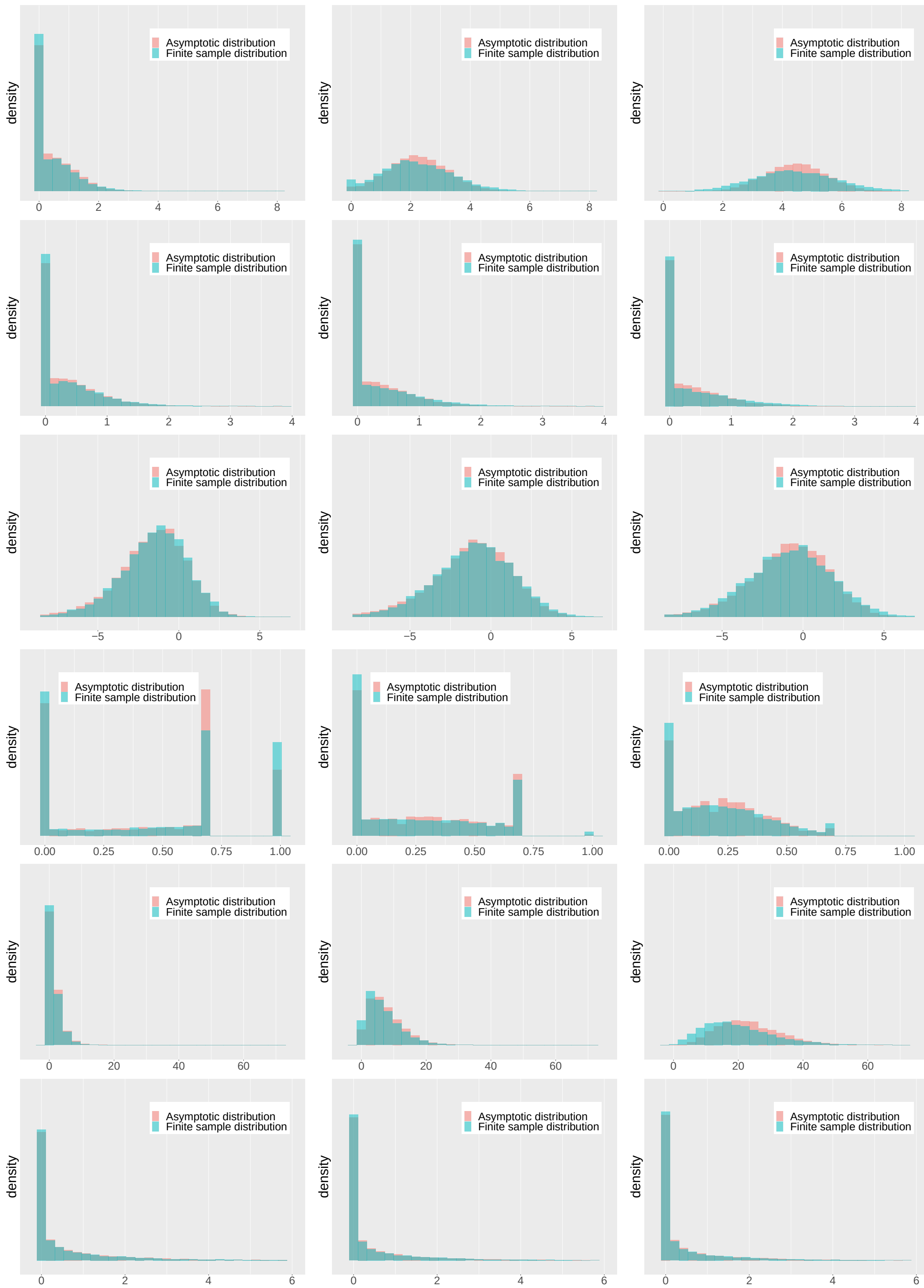
$$\begin{pmatrix} z_t \\ \epsilon_t \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{z\epsilon} \\ \rho_{z\epsilon} & 1 \end{pmatrix} \right).$$

In what follows, we vary ρ_x and $\rho_{z\epsilon}$. Note that, for this *dgp*, z_t and x_t are independent if and only if $\rho_{z\epsilon} = 0$.

4.1 $n = 500,\ \rho_{z\epsilon} = 0,\ \rho_x = 0.5,\ \bar{\pi} = 0.7,\ \pi = 0.2$

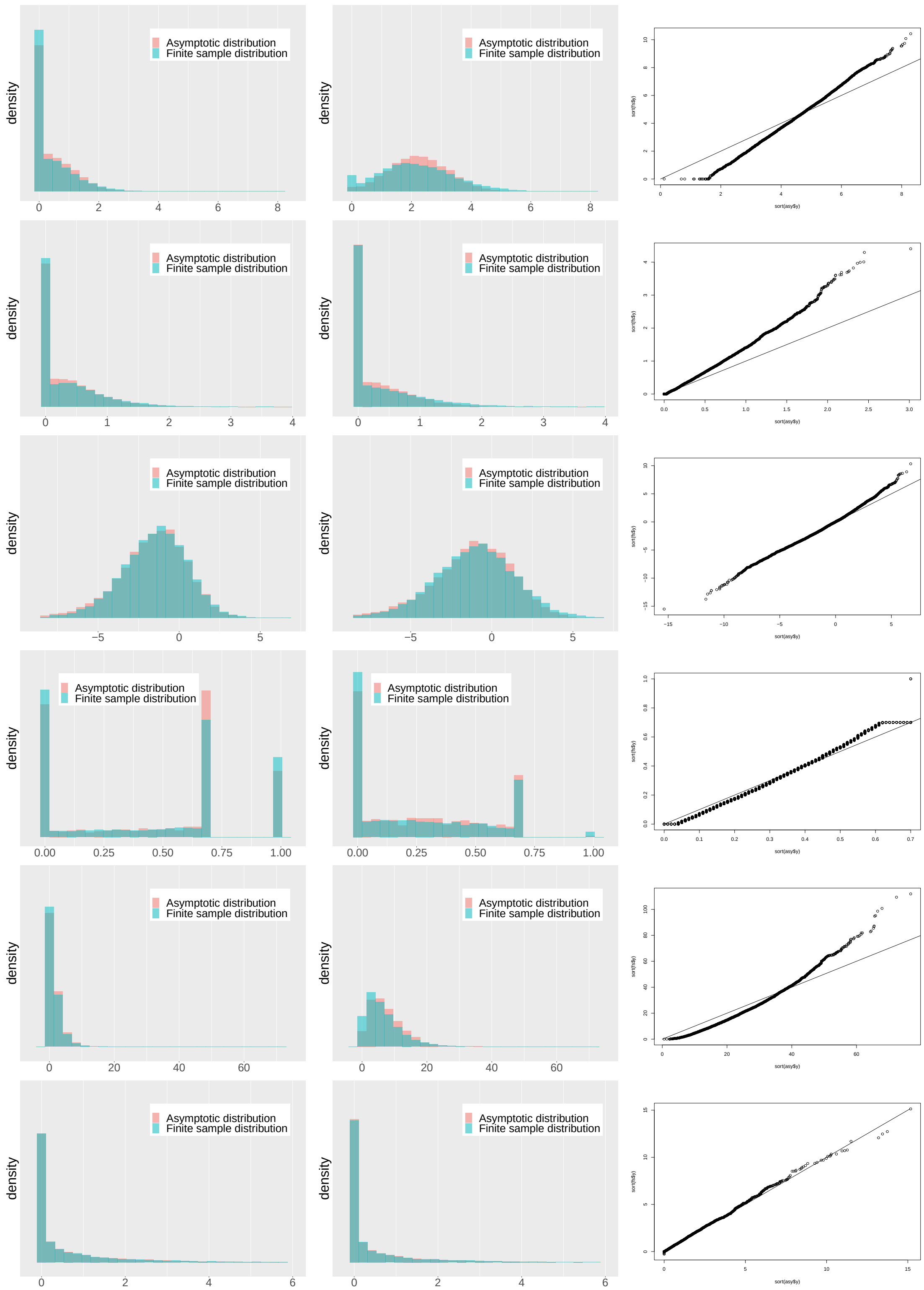


4.2 $n = 1000,\ \rho_{z\epsilon} = 0,\ \rho_x = 0.5,\ \bar{\pi} = 0.7,\ \pi = 0.2$

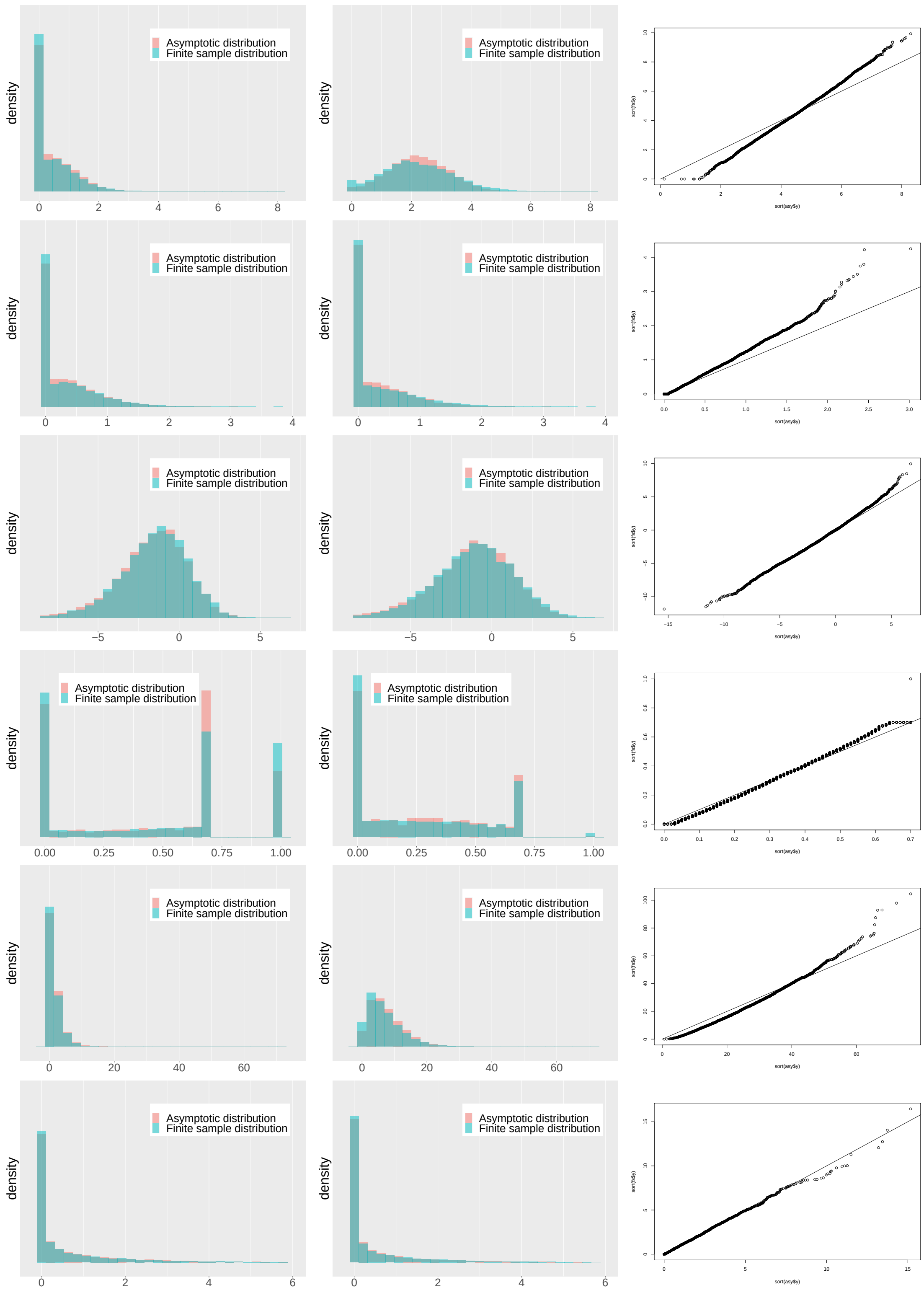


4.3 $n = 10000,\ \rho_{z\epsilon} = 0,\ \rho_x = 0.5,\ \bar{\pi} = 0.7,\ \pi = 0.2$

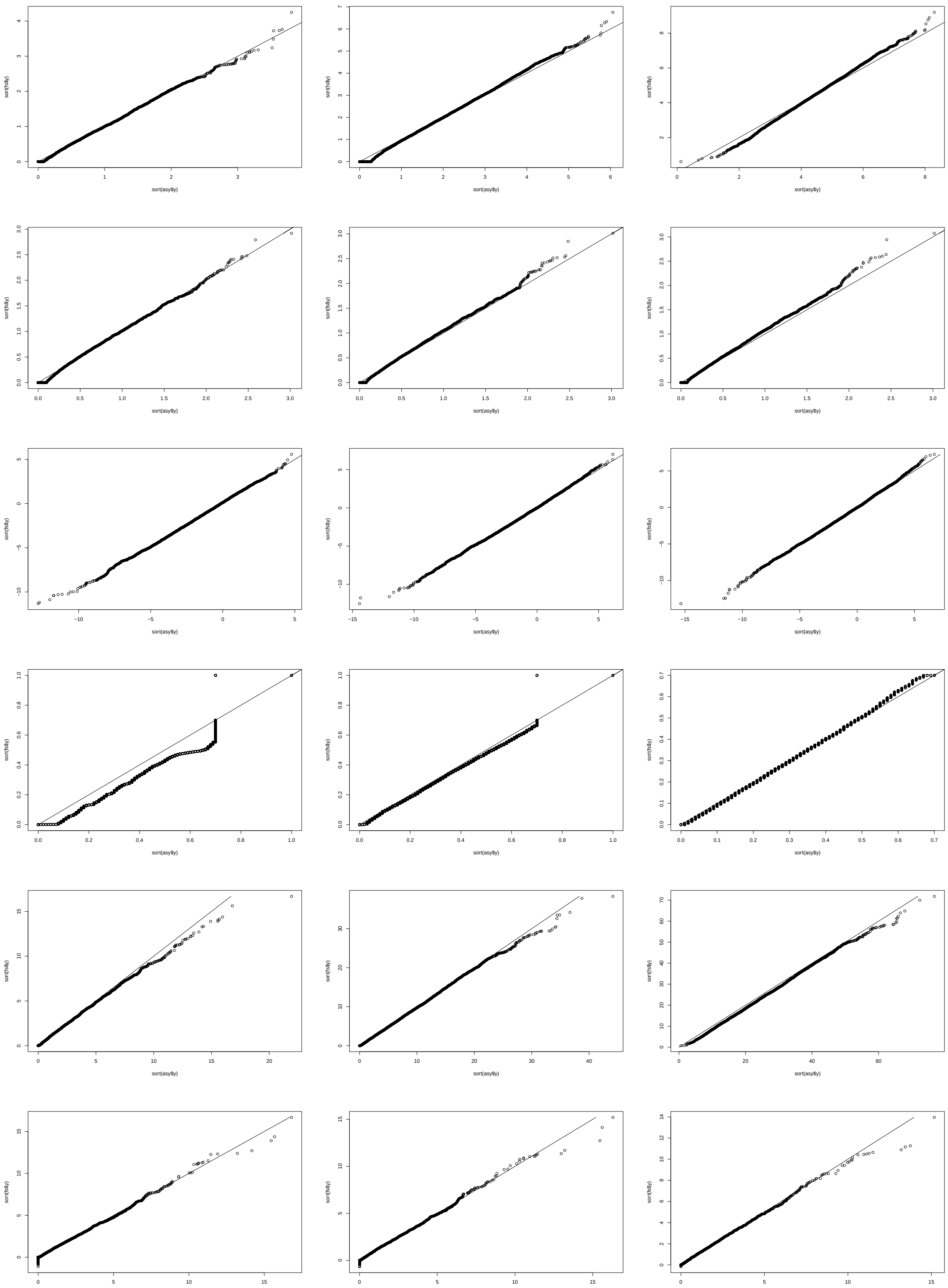
4.4 $n = 500,\ \rho_{z\epsilon} = 0,\ \rho_x = 0.5,\ \bar{\pi} = 0.7,\ \pi = 0.2$



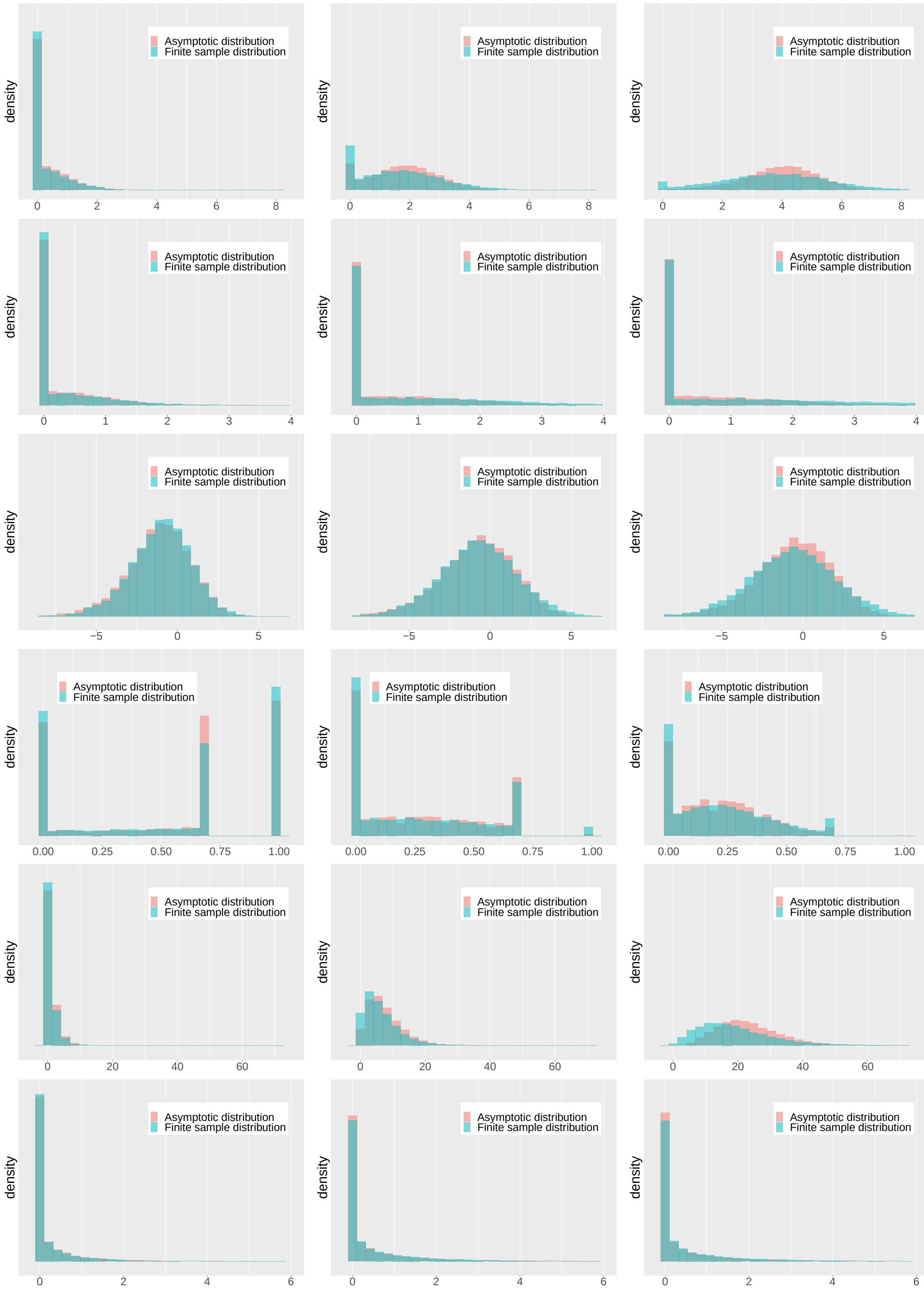
4.5 $n = 1000,\ \rho_{z\epsilon} = 0,\ \rho_x = 0.5,\ \bar{\pi} = 0.7,\ \pi = 0.2$



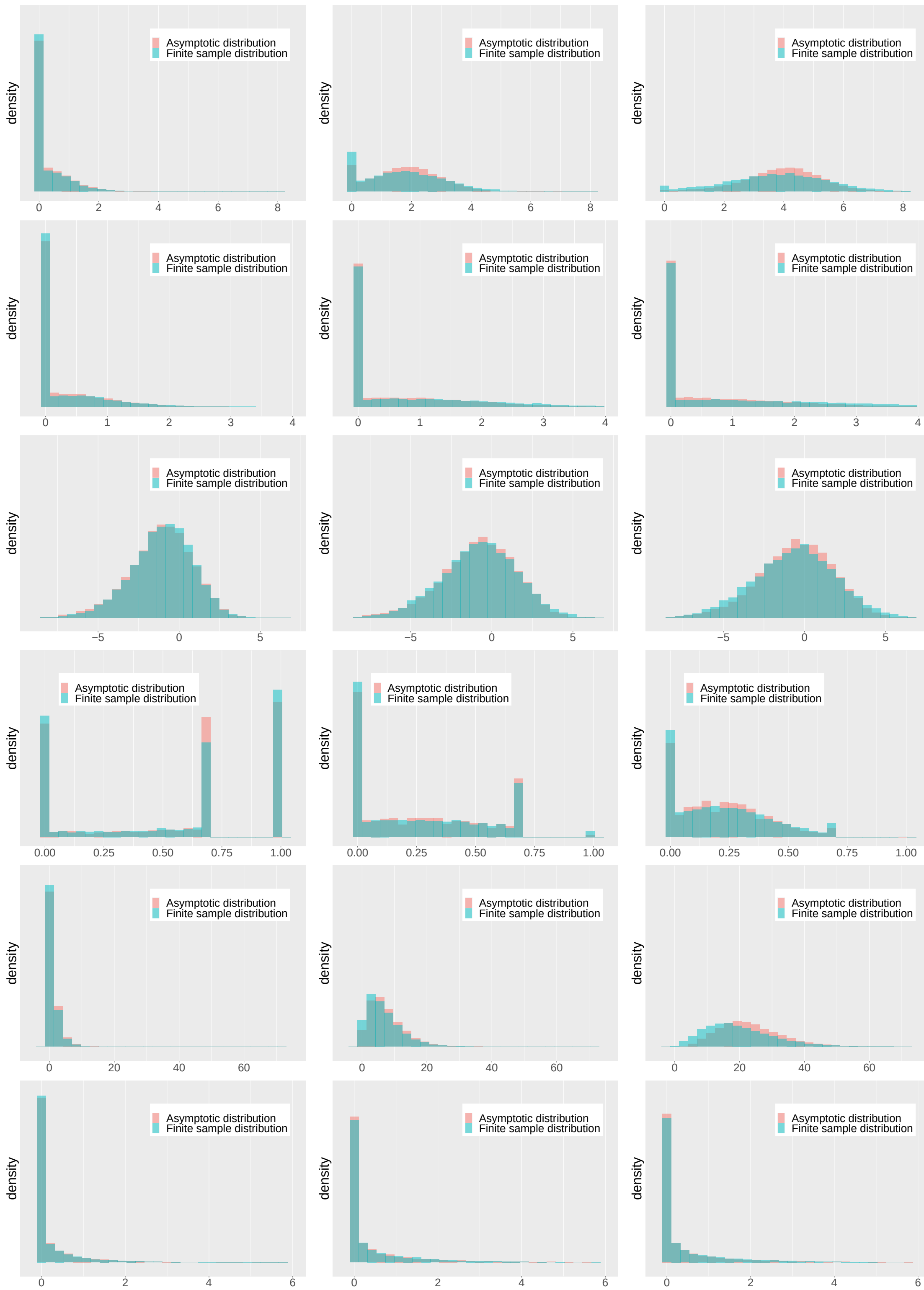
4.6 $n = 10000$, $\rho_{z\epsilon} = 0$, $\rho_x = 0.5$, $\bar{\pi} = 0.7$, $\pi = 0.2$



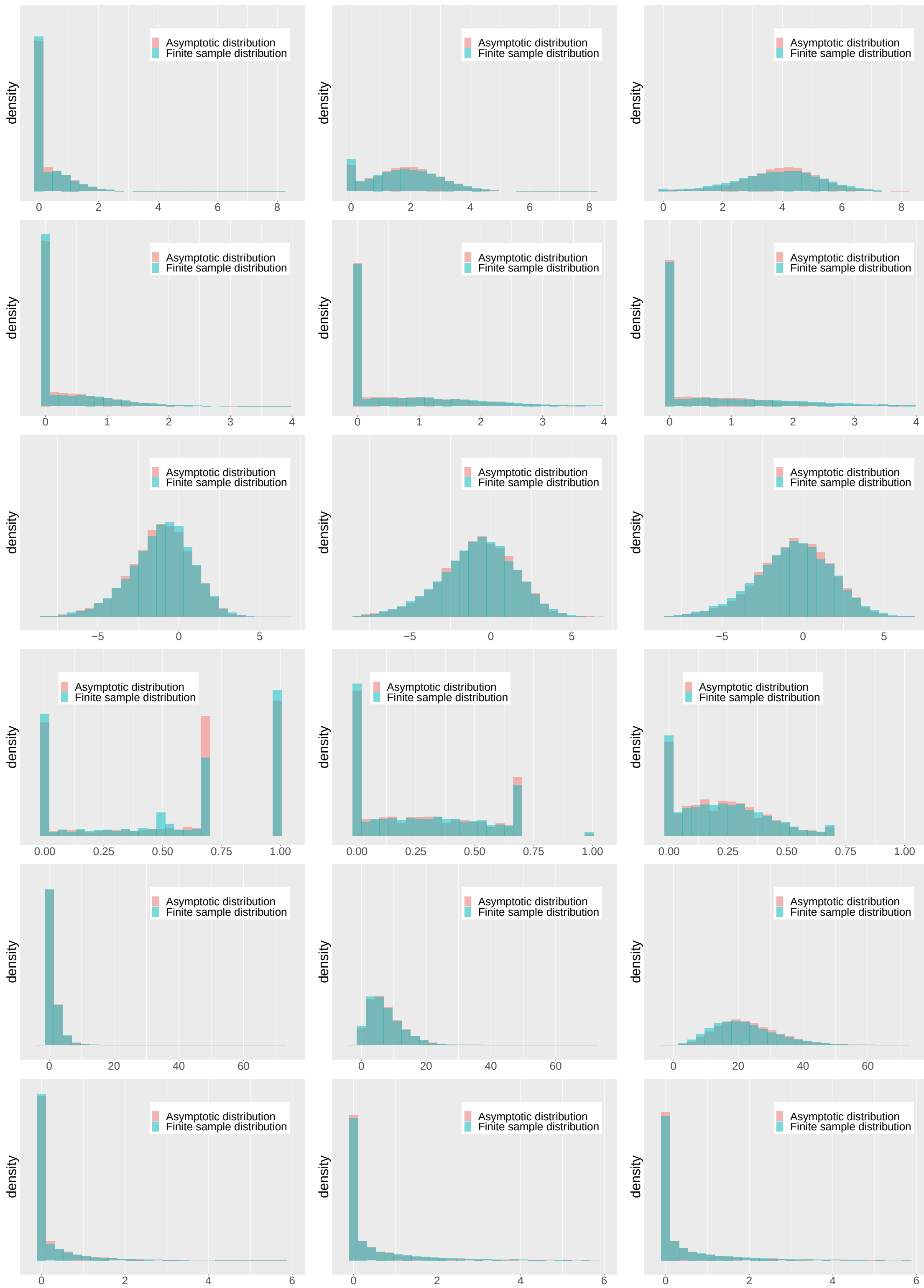
4.7 $n = 500,\ \rho_{z\epsilon} = 0.9,\ \rho_x = 0,\ \bar{\pi} = 0.7,\ \pi = 0.2$



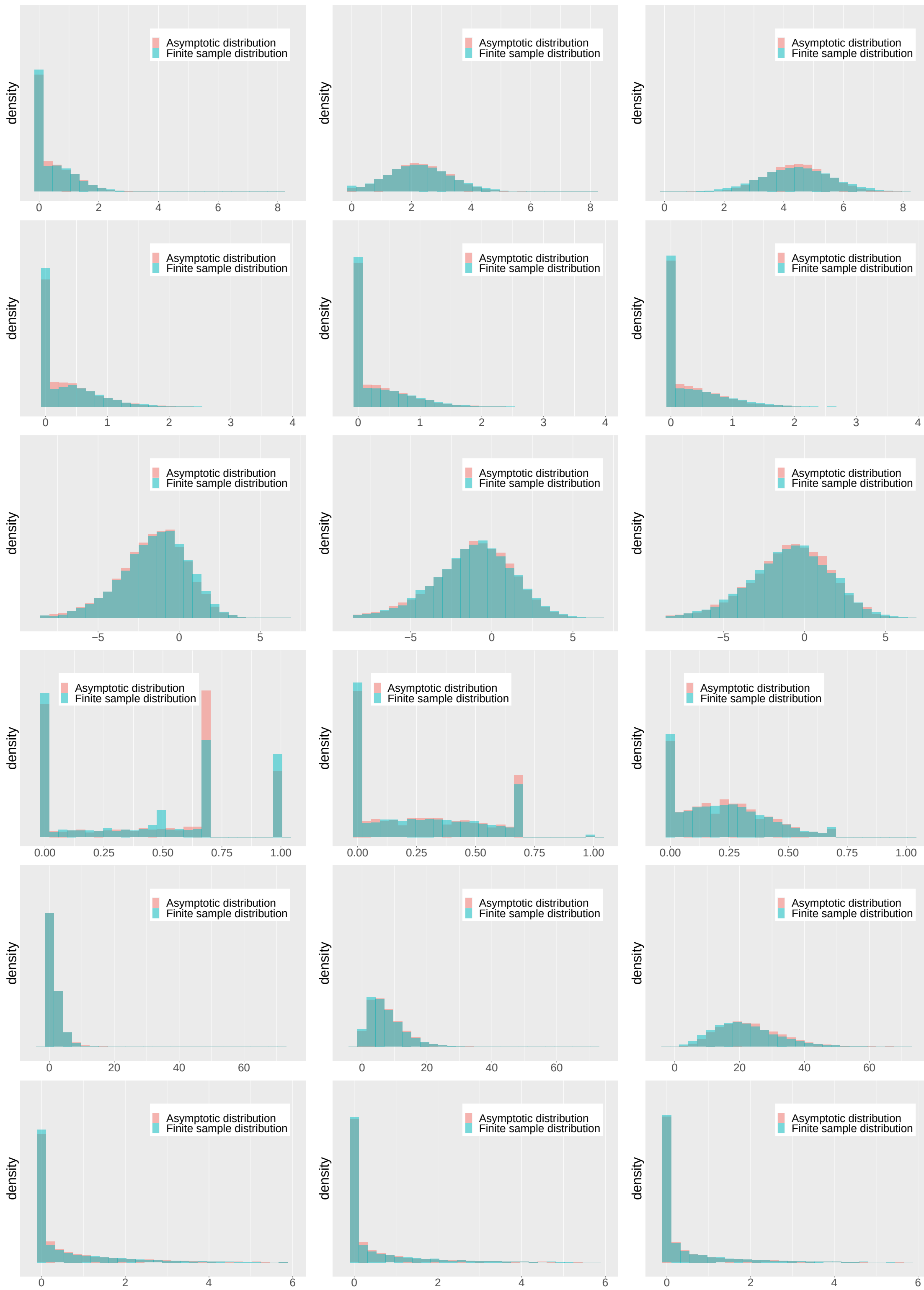
4.8 $n = 1000,\ \rho_{z\epsilon} = 0.9,\ \rho_x = 0,\ \bar{\pi} = 0.7,\ \pi = 0.2$



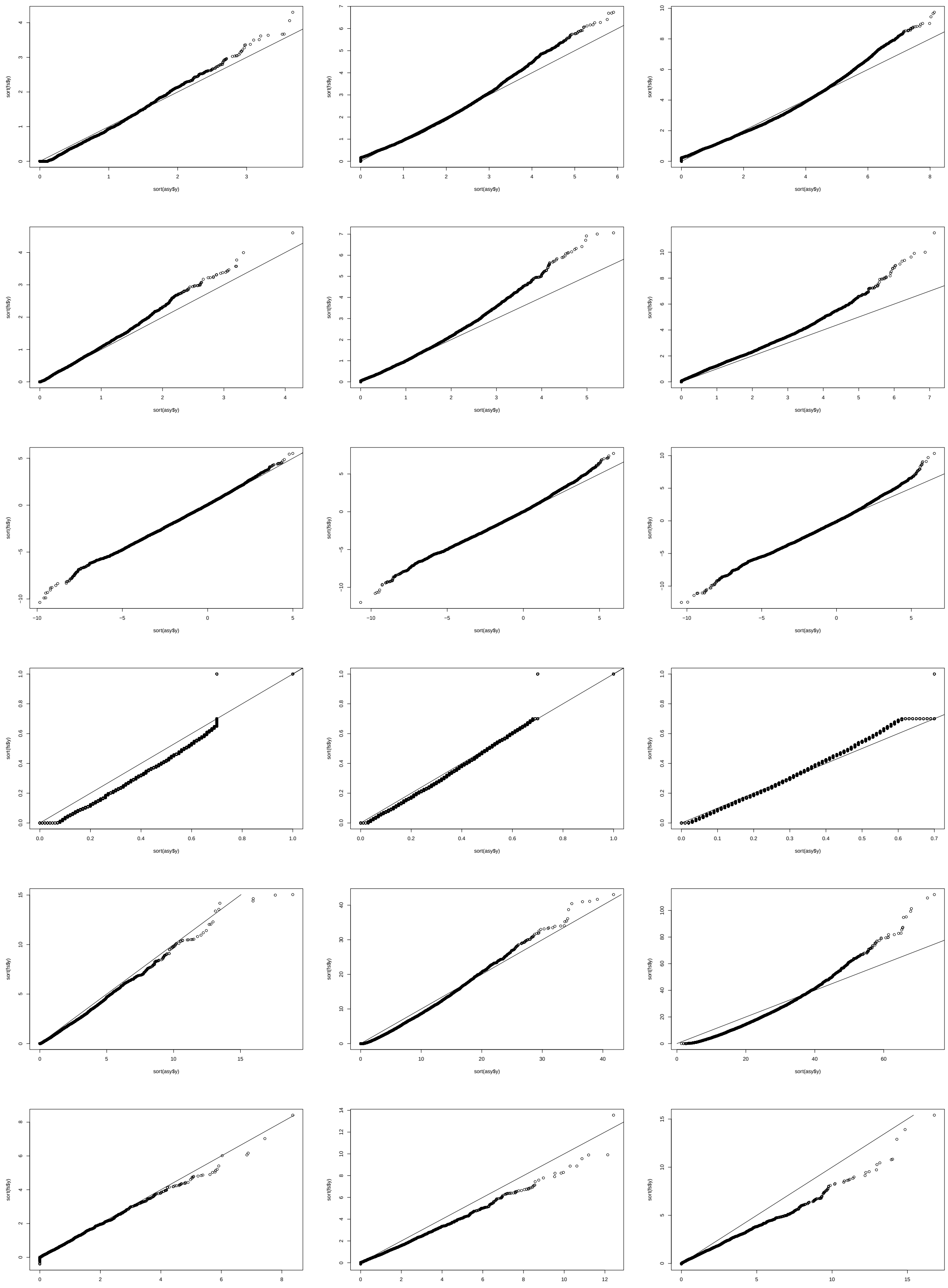
4.9 $n = 10000,\ \rho_{z\epsilon} = 0.9,\ \rho_x = 0,\ \bar{\pi} = 0.7,\ \pi = 0.2$



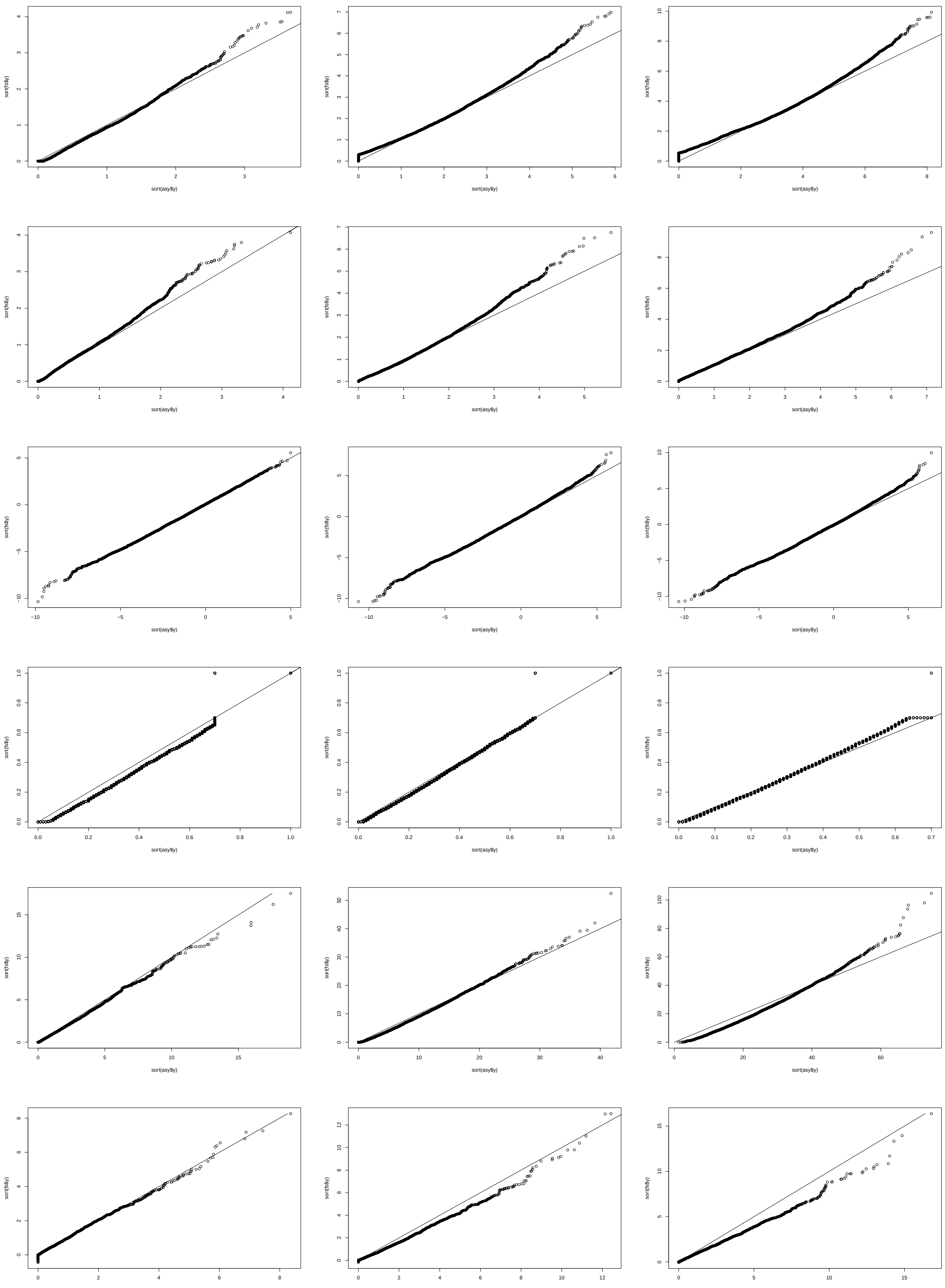
4.10 $n = 10000,\ \rho_{z\epsilon} = 0,\ \rho_x = 0.5,\ \bar{\pi} = 0.7,\ \pi = 0.2$



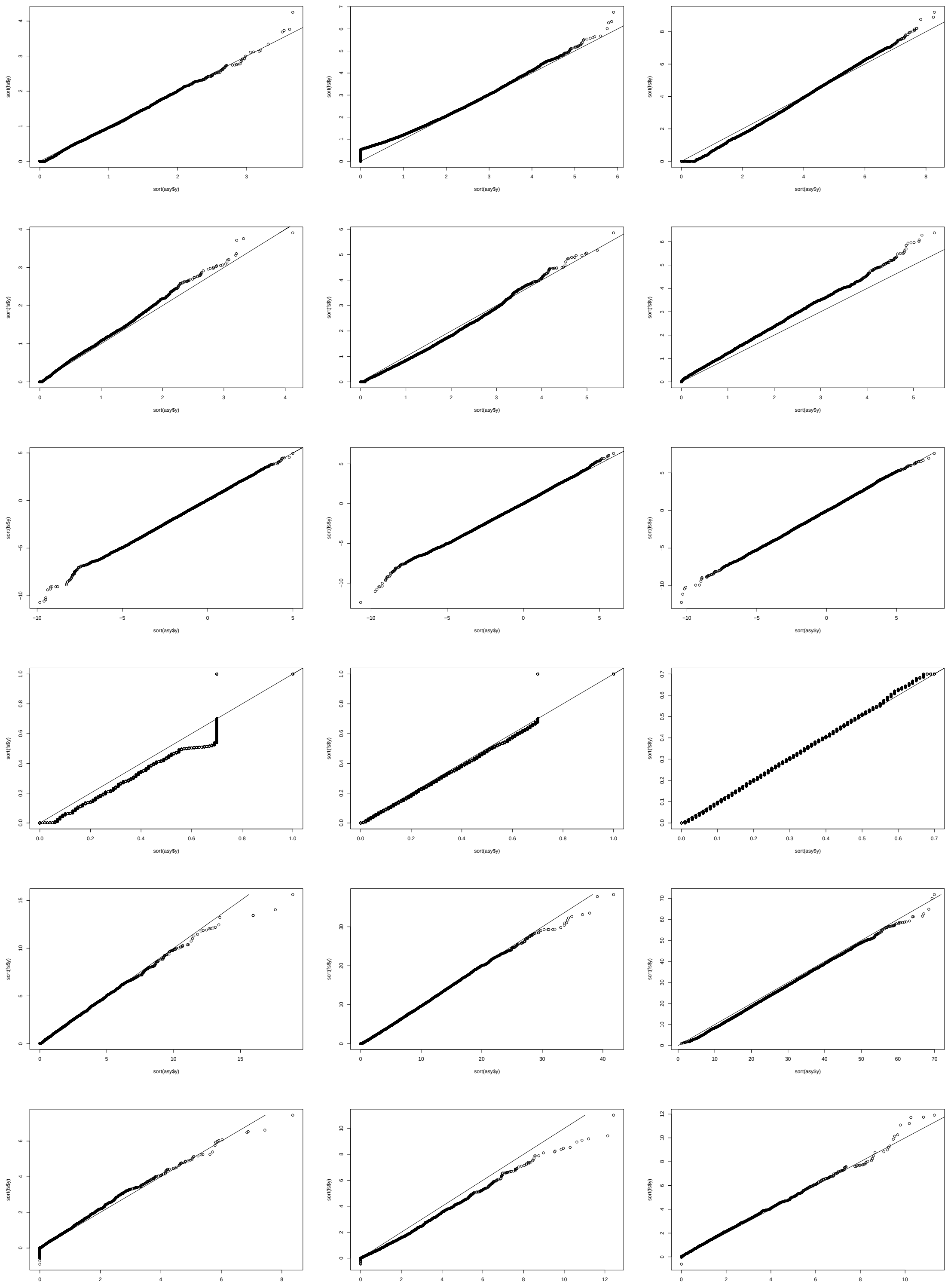
4.11 $n = 500,\ \rho_{z\epsilon} = 0.9,\ \rho_x = 0,\ \bar{\pi} = 0.7,\ \pi = 0.2$



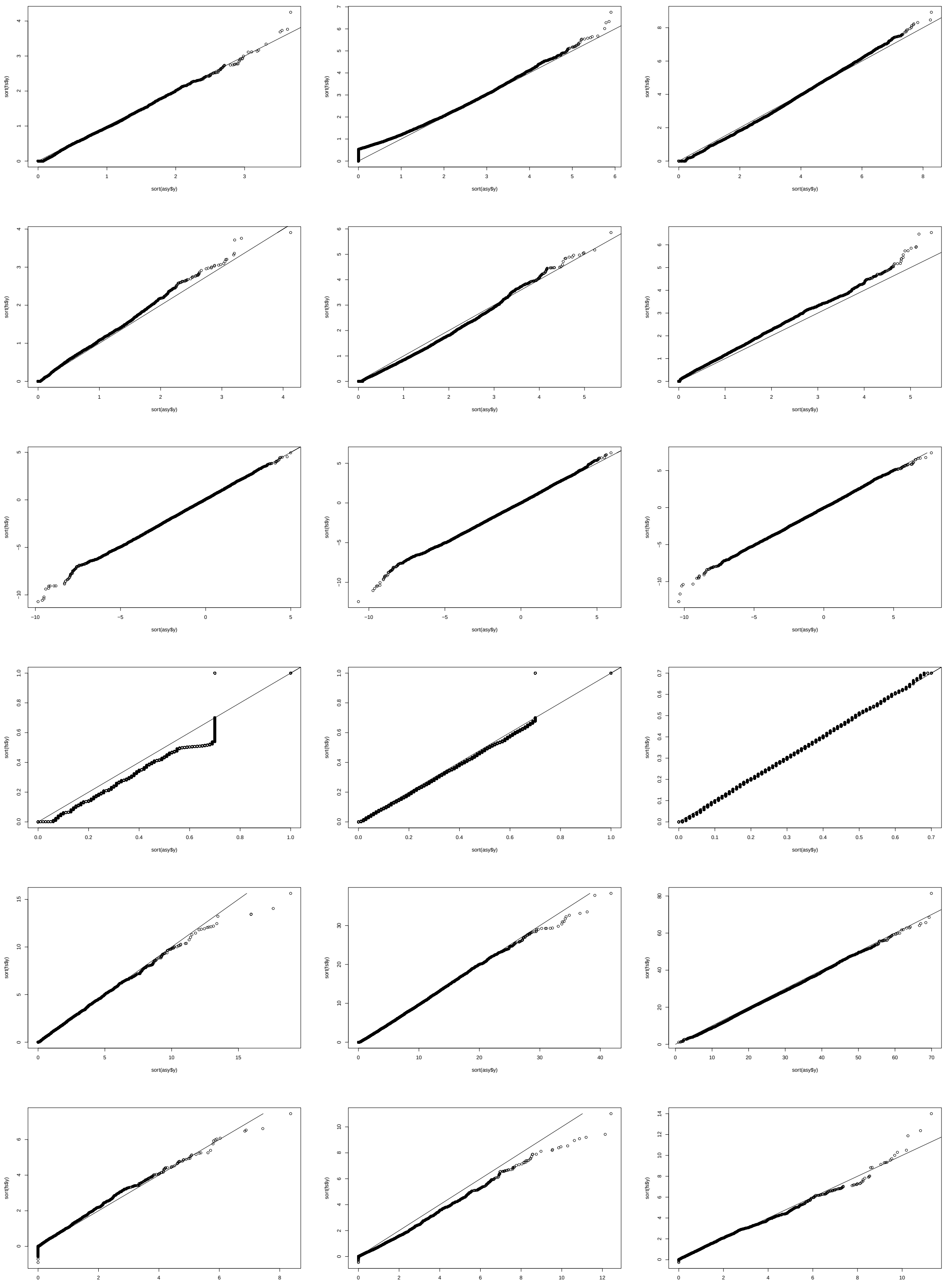
4.12 $n = 1000$, $\rho_{z\epsilon} = 0.9$, $\rho_x = 0$, $\bar{\pi} = 0.7$, $\pi = 0.2$



4.13 $n = 10000$, $\rho_{z\epsilon} = 0.9$, $\rho_x = 0$, $\bar{\pi} = 0.7$, $\pi = 0.2$



4.14 $n = 20000$, $\rho_{z\epsilon} = 0.9$, $\rho_x = 0$, $\bar{\pi} = 0.7$, $\pi = 0.2$



5 Power comparison

[To do: Complete the subsequent "preliminary" simulation results.] We choose $\overline{\pi} = 0.7$. The true parameters are set equal to $\beta_2 = 0$, $\zeta = 1$, $\pi = 0.2$. We set $n = 500$, with a burn-in phase of 100 observations. In the simulation results below, which are based on 10,000 replications (for both, asymptotic and finite sample distributions), we vary β_1 in the set $\{0, 0.1, 0.2\}$ which correspond to the localization parameter $b_1 = \sqrt{500} \cdot \beta_1 \in \{0, 2.23, 4.47\}$. x_t is drawn from an AR(1), i.e.,

$$x_t = \rho x_{t-1} + \epsilon_t.$$

We set $\rho = 0.5$ and generate data from $z_t, \epsilon_t \overset{iid}{\sim} N(0, 1)$.

In Figures **??**–**??**, the scale *within* each figure is the same, but this is not true *between* figures.

Figures **??**–**??** show that the asymptotic and finite sample distributions of $\hat{\beta}_{n,1}, \hat{\beta}_{n,2}, \hat{\zeta}_n$, and $\hat{\pi}_n$ (appropriately demeaned, or not, and scaled) for $\beta_1 \in \{0, 0.1, 0.2\}$ from left to right. Note that for both, the asymptotic and the finite sample distributions, $\hat{\pi}_n$ and its asymptotic counterpart are set equal to 1 whenever they are not defined, which happens when $\hat{\beta}_n$ or its asymptotic counterpart are equal to $(0, 0)'$. The take-away messages from Figures **??**–**??** are:

- The asymptotic approximations work well. [To do: There is some room for improvement by replacing $\Omega(\pi_1, \pi_2; \gamma_0)$ with the covariance kernel evaluated at the true parameter (i.e., $\beta_1 \in \{0, 0.1, 0.2\}$ rather than $\beta_1 = 0$ throughout). This should somewhat increase the asymptotic variance, which is in line with the graphs where the asymptotic distributions currently seem too tight.]

- While $\beta_1 = 0.2$ is far enough from the boundary for the corresponding estimator, $\hat{\beta}_{n,1}$, to no longer take on the value 0 (cf. Figure **??**), weak identification is still persuasive (unclear in how far this distribution is well approximated by a truncated normal), cf. Figure **??**.

Figure **??** shows the asymptotic and finite sample distribution of the LR statistic for testing $H_0 : \beta_2 = 0$ for $\beta_1 \in \{0, 0.1, 0.2\}$ from left to right. Since $\beta_2 = 0$ throughout, all distributions are "null distributions."

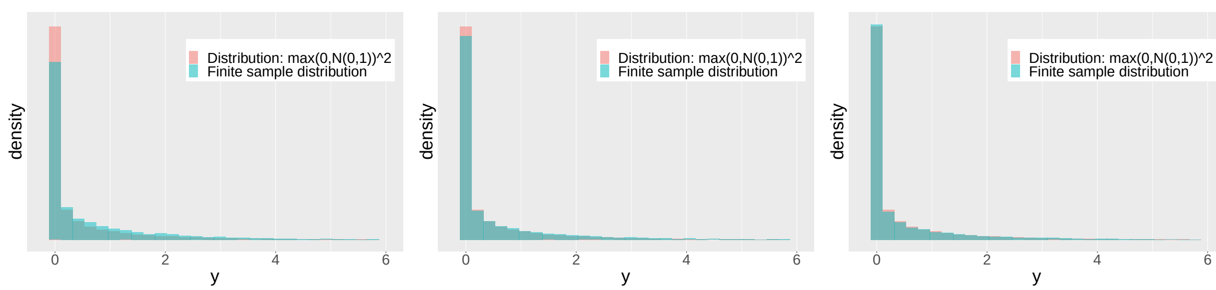


Figure 1: Histogram of $\max(0, N(0,1))^2$ and finite sample distribution ($n = 500$) of LR statistic for testing $H_0 : \beta_2 = 0$, with $b_1 = \sqrt{500}\cdot 0\ (= 0)$, $\sqrt{500}\cdot 0.1\ (\approx 2.23)$, and $\sqrt{500}\cdot 0.2\ (\approx 4.47)$ from left to right.

Because the three graphs in Figure **??** look somewhat similar, Figure 1 also plots the finite sample distribution of the LR statistic, but compares it to the asymptotic distribution that Pedersen and Rahbek (2019) obtain for $\beta_1 \gg 0$ ($b_1 = \infty$), i.e., $\max(0, N(0,1))^2$. The right graph shows that even though $\beta_1 = 0.2$ ($b \approx 4.47$) is not super far from the boundary, the asymptotic distribution result in Pedersen and Rahbek (2019) seems to already kick in. (This is of course no uniformly valid statement, because everything depends on the distribution of x_t as well as the true value of π.) Plus, the left graph shows that even when $\beta_1 = 0$, the $\max(0, N(0,1))^2$ is not too far off (see Section 5.1 for the expression of the asymptotic distribution). This already hints at the potential power gains.

Figure **??** shows the asymptotic and finite sample distribution of the LR^* statistic for testing $H_0^* : \beta_1 = \beta_2 = 0$ for $\beta_1 \in \{0, 0.1, 0.2\}$ from left to right. [While the approximation by the asymptotic distribution doesn't look that good here, note the scale on the x-axis, i.e., the shift in the mean is still well captured. Similar to above, the asymptotic approximation can be improved upon.] The left hand graph, thus, shows distributions under the null, while the other two graphs show distributions under the alternative. By "combining" the graphs in Figures **??** and **??**, we obtain the asymptotic null rejection probability of the two step testing procedure. [To do: Add the corresponding simulation results.]

5.1 Analysis of sequential test

[To do: Present the power comparison of the two step testing procedures and the "second step" test (or LR) statistic with plug-in least favorable configuration critical values.]

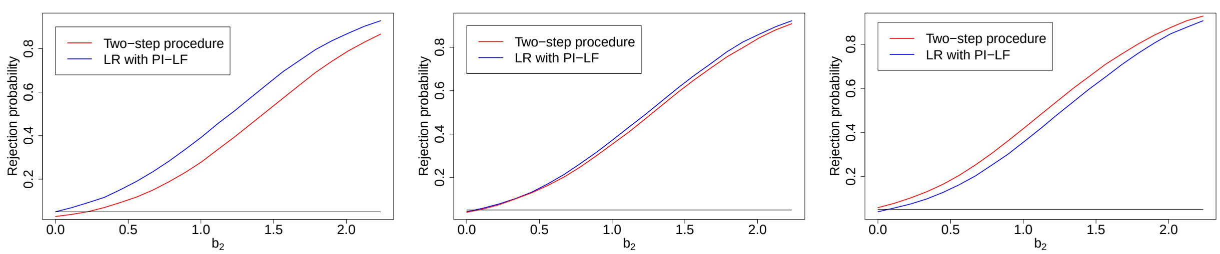


Figure 2: Asymptotic power as a function of b_2, with $b_1 = \sqrt{500} \cdot 0$ $(= 0)$, $\sqrt{500} \cdot 0.05$ (≈ 1.12), and $\sqrt{500} \cdot 0.1$ (≈ 2.24) from left to right. LR with PI-LF in black. Two-step procedure in red. $\pi = 0.2$.

6 Asymptotic size of two-step procedures

[To do!]

7 Conclusion

[To do!]

References

Andrews, D. W. K. (2001). Testing When a Parameter Is on the Boundary of the Maintained Hypothesis. *Econometrica 69*(3), 683–734.

Andrews, D. W. K. and X. Cheng (2012). Estimation and Inference With Weak, Semi-Strong, and Strong Identification. *Econometrica 80*(5), 2153–2211.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*(1), 21–59.

Leeb, H. and B. M. Pötscher (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory 24*(2), 338–376.

Pedersen, R. S. and A. Rahbek (2019). Testing GARCH-X Type Models. *Econometric Theory 35*(5), 1012–1047.