Applications of Copula Theory and Regime Switching in Finance

Alex Donov

A thesis submitted for the degree of Doctor of Philosophy

University of Essex

Abstract

There is well-documented evidence that the dependence structure of financial assets is often characterized by considerable time variation. Financial markets are repeatedly subjected to episodes of rapid growth and dramatic decline of asset prices, and the recent financial crisis reinforced the need to model extreme events and sudden changes in the behaviour of financial assets. In particular, financial returns have been shown to exhibit stronger tail dependence during financial downturn. That is, extreme negative events are highly correlated and tend to cluster together. Traditional static models, such as the multivariate Normal distribution, are unable to capture these characteristics of the dependence structure, which resulted in copula models attracting attention and becoming popular over the last decade. Copulas provide greater flexibility by allowing the dependence structure to be modelled separately from marginal distributions. Furthermore, a rich class of higher dimensional copulas with various types of asymmetric tail dependence can be constructed through the use of vine copulas.

The objective of this research work is take into account the time-varying dependence structure by combining copula theory with regime switching models that exhibit Markov property. In this class of models the dependence structure is assumed to switch between regimes according to a hidden state variable, with the purpose of accurately describing the behaviour of financial time series. Furthermore, the goal is to extend this class of models to higher dimensions where complex dependence characteristics are also present. Applications of these models are not restricted to finance and can be useful in any context where the dependence structure amongst random variables changes over time.

Contents

List of Figures

List of Tables

Introduction

There is well-documented evidence that the dependence structure of financial assets is often characterized by considerable time variation. Financial markets are repeatedly subjected to episodes of rapid growth and dramatic decline of asset prices, and the recent financial crisis reinforced the need to model extreme events and sudden changes in the behaviour of financial assets. In particular, financial returns have been shown to exhibit stronger tail dependence during financial downturn. That is, extreme negative events are highly correlated and tend to cluster together. Traditional static models, such as the multivariate Normal distribution, are unable to capture these characteristics of the dependence structure, which resulted in copula models attracting attention and becoming popular over the last decade. Copulas provide greater flexibility by allowing the dependence structure to be modelled separately from marginal distributions. Furthermore, a rich class of higher dimensional copulas with various types of asymmetric tail dependence can be constructed through the use of vine copulas.

The objective of this research work is take into account the time-varying dependence structure by combining copula theory with regime switching models that exhibit Markov property. In this class of models the dependence structure is assumed to switch between regimes according to a hidden state variable.

In **Chapter 1** we employ Silva Filho et al. (2012) specification to model the dynamic dependence structure between S&P500 and FTSE100 stock indices. The dependence parameter of the chosen copulas evolves according to Patton (2006) framework, whilst the intercept switches according to a first-order Markov chain. This modelling approach identifies two distinct regimes: one of high dependence and one of low dependence. Using symmetrized Joe–Clayton copula we find that the low dependence regime is characterised by asymmetric dependence, whilst mild asymmetry is detected in the high dependence regime. We compute standard errors for parameters in our models using standardized residuals bootstrap procedure.

In **Chapter 2** we model the dependence structure in international financial returns using second-order regime switching vine copula. Gaussian copula is used as a main building

block in the vine specification where the dependence parameters are held constant within each regime. The model is applied to returns of the S&P500, FTSE100 and DAX stock indices. The standard errors of the estimates are computed using Godambe information matrix. Empirically, regimes of high dependence are identified. The model is then compared against the benchmark model in which regime variable follows first-order process. Information criteria such as AIC and BIC suggest that the second-order regime switching model may be a better choice.

In **Chapter 3** we examine the conditions under which models of this variety work well. The sort of environment we are considering is where regime-switching copula is applied to situations such as recessions, financial crises, all sorts of unusual events where one would believe that a particular state is much less likely to occur than the other. For example, if the government regulatory policy becomes more effective, one would expect crises to be rare over time. The concern could be that applying Markov-switching models in situations where one regime occurs only very rarely may be problematic since one is not in that regime very frequently. Subsequently, the amount of information one obtains from that regime is going to be small, and is likely that the characteristics of that regime will be poorly determined. As a result, the usual asymptotic theory may not be particularly reliable with regards to the parameters relevant to that regime, namely copula and transition parameters. This is because the standard asymptotic theory is based on the assumption that the probability transition matrix remains fixed as the sample size increases to infinity. Hence, the asymptotic sequence in which the fraction of the time spent in one of the states converges to something moderate or zero may not be informative about the finite sample distributions. In this situation, the asymptotic theory that is generally used would not be appropriate. The goal of this chapter is to provide a mathematical framework to analyse scenarios under the alternative asymptotic sequence, and conduct an investigation through a Monte Carlo study in order to examine finite sample properties of a Maximum Likelihood Estimator under the alternative asymptotic sequence.

Preliminaries

1 Copula Theory

Let $U_1, \ldots, U_d$ be independent random variables defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Suppose that each U_i is uniformly distributed on $[0,1]$. A d-dimensional copula $C(u_1, ..., u_d)$ is a multivariate distribution function in $[0,1]^d$ whose marginal distribution are uniform in [0,1] interval. A d-dimensional copula has the following properties:

1. $C(u_1, \ldots, u_j, \ldots, u_d) = 0$ if $u_j = 0$ for at least one $j \in \{1, \ldots, d\}$
2. $C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$ for all u_j and $j \in \{1, \ldots, d\}$
3. C is d-increasing, that is, for all $\boldsymbol{a} = (a_1, \ldots, a_d) \in [0,1]^d$ and $\boldsymbol{b} = (b_1, \ldots, b_d) \in [0,1]^d$, where $a_i \leqslant b_i$:

$$\Delta_{(\boldsymbol{a},\boldsymbol{b}]} C = \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{\sum_{j=1}^{d} i_j} C\left(u_{1,i_1}, \ldots, u_{d,i_d}\right) \geqslant 0 \tag{1}$$

where $u_{j,1} = a_j$ and $u_{j,2} = b_j$ for all $j \in \{1, \ldots, d\}$

Sklar's Theorem: If $X_1, ..., X_d$ has joint cumulative distribution function $F_{1,...,d}(x_1, ..., x_d)$ and marginal cumulative distribution functions $F_1(x_1), ..., F_d(x_d)$, then there exists appropriate copula C such that for all $x_1, ..., x_d$ in $\bar{R}$:

$$F_{1,...,d}(x_1, ..., x_d) = C_{1,...,d}\left(F_1(x_1), ..., F_d(x_d)\right) \tag{2}$$

When $F_1(x_1), ..., F_d(x_d)$ are continuous, then $C_{1,...,d}$ is unique. All $U_1 = F_1(X_1), ..., U_d = F_d(X_d)$ have $U(0,1)$ distributions, and $C_{1,..,d}(u_1, ..., u_d)$ is the joint cumulative distri-

bution function of $(U_1, ..., U_d)$. The joint probability density function $f_{1,...,k}(x_1, ..., x_d)$ for an absolutely continuous $F_{1,...,d}(x_1, ..., x_d)$ with strictly increasing continuous margins $F_1(x_1), ..., F_d(x_d)$ is:

$$f_{1,...,k}(x_1, ..., x_d) = c_{1,..,d}(u_1, ..., u_d) \prod_{i=1}^{d} f_i(x_i) \tag{3}$$

where $c_{1,..,d}(u_1, ..., u_d) = \frac{\partial^d C_{1,...,d}(u_1,...,u_d)}{\partial u_1 ... \partial u_d}$.

1.1 Conditional Copula

Patton (2006) extended theorem of Sklar (1959) to conditional distributions.

Theorem 1. *Let $F_{X|W}(\cdot|w)$ be the conditional distribution of $X|W = w$, $F_{Y|W}(\cdot|w)$ be the conditional distribution of $Y|W = w$, $F_{XY|W}(\cdot|w)$ be the joint conditional distribution of $X, Y|W = w$, and $\mathcal{W}$ be the support of W. Assuming that $F_{X|W}(\cdot|w)$ and $F_{Y|W}(\cdot|w)$ are continuous in x and y for all $x \in \mathcal{W}$, there exist a unique conditional copula $C(\cdot|w)$ such that:*

$$F_{XY|W}(x, y|w) = C(F_{X|W}(x|w)), F_{Y|W}(y|w)|w) \tag{4}$$

In this extension of Sklar's theorem it is important that the conditional variable, W, is the same for both marginal distributions and the copula. If this condition is not met, then $F_{XY|W}(x, y|w)$ will fail to satisfy the condition for the joint conditional distribution.

2 The models for copula

There exist many functional forms that can be used as copulas. In this section we present functional forms of copulas that will be used in subsequent chapters.

2.1 Gaussian (Normal) Copula

Normal copula has the linear correlation coefficient ρ as its dependence parameter, although it has no tail dependence.

$$C_G(u_1, u_2|\rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{-(r^2 - 2\rho rs + s^2)}{2(1-\rho^2)}\right\} drds$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal, and $\rho \in (-1, 1)$.

2.2 Student (Student-t) copula

Student-t copula has also the linear correlation coefficient ρ as a measure of dependence. Although Student-t copula has tail dependence, it imposes symmetry in both tails.

$$C_t(u_1, u_2|\rho, \nu) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{r^2 - 2\rho rs + s^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} drds$$

where ν is the degree-of-freedom parameter, $t_\nu^{-1}(\cdot)$ is the inverse of the standard Student-t cumulative distribution function, and $\rho \in (-1, 1)$.

2.3 Symmetrised Joe-Clayton copula (SJC)

Symmetrised Joe-Clayton copula (SJC) was proposed by Patton (2006, p. 542) and is given by:

$$C_{SJC}(u_1, u_2|\tau_U, \tau_L) = \frac{1}{2} C_{JC}(u_1, u_2|\tau_U, \tau_L) + \frac{1}{2}\left(C_{JC}(1-u_1, 1-u_2|\tau_U, \tau_L) + u_1 + u_2 - 1\right)$$

where $\tau_L \in (0, 1)$ and $\tau_U \in (0, 1)$ are the coefficients of lower and upper tail dependence respectively, and C_{JC} is the unmodified Joe-Clayton, which is also known as BB7 copula.

2.4 Gumbel copula

The distribution of Gumbel copula has the following form:

$$C_G(u_1, u_2|\theta) = \exp\left(-\left[(-\log u_1)^\theta + (-\log u_2)^\theta\right]^{\frac{1}{\theta}}\right)$$

2.5 Clayton copula

The distribution of Clayton copula has the following form:

$$C_C(u_1, u_2|\theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}$$

2.6 Frank copula

The distribution of Frank copula has the following form:

$$C_F(u_1, u_2|\theta) = -\delta^{-1}\log([(1 - e^{-\delta}) - (1 - e^{-\delta u_1})(1 - e^{-\delta u_2})]/1 - e^{-\delta})$$

2.7 Joe copula

The distribution of Joe copula has the following form:

$$C_J(u_1, u_2|\theta) = 1 - ((1 - u_1)^{\delta} + (1 - u_2)^{\delta} - (1 - u_1)^{\delta}(1 - u_2)^{\delta})^{1/\delta}$$

2.8 BB1 copula

The distribution of BB1 copula has the following form:

$$C_{BB1}(u_1, u_2|\theta, \delta) = \left(1 + [(u_1^{-\theta} - 1)^{\delta} + (u_2^{-\theta} - 1)^{\delta}]^{\frac{1}{\delta}}\right)^{-\frac{1}{\theta}}$$

where $\theta \geq 1, \delta > 0$.

2.9 BB6 copula

The distribution of BB6 copula has the following form:

$$C_{BB6}(u_1, u_2|\theta, \delta) = 1 - \left(1 - \exp\left\{-[(-\log(1 - \tilde{u}_1^{\theta}))^{\delta} + (-\log(1 - \tilde{u}_2^{\theta}))^{\delta}]^{\frac{1}{\delta}})^{\frac{1}{\theta}}\right\}\right)$$

where $\tilde{u}_1 = 1 - u_1$, $\tilde{u}_2 = 1 - u_2$, and $\theta \geq 1, \delta \geq 1$.

2.10 BB7 copula

The distribution of BB7 copula has the following form:

$$C_{BB7}(u_1, u_2|\theta, \delta) = 1 - \left(1 - [(1-\tilde{u}_1^{\theta})^{-\delta} + (1-\tilde{u}_2^{\theta})^{-\delta} - 1]^{-\frac{1}{\delta}}\right)^{\frac{1}{\theta}}$$

where $\tilde{u}_1 = 1 - u_1$, $\tilde{u}_2 = 1 - u_2$. Given tail parameters τ_U and τ_L: $\theta = \frac{1}{log_2(2-\tau_U)} \geq 1$, and $\delta = \frac{-1}{log_2(\tau_L)} > 0$.

2.11 BB8 copula

The distribution of BB8 copula has the following form:

$$C_{BB8}(u_1, u_2|\theta, \delta) = \delta^{-1}[1 - \left\{1 - [1-(1-\delta)^{\theta}]^{-1}[1-(1-\delta u_1)^{\theta}][1-(1-\delta u_2)^{\theta}]\right\}^{\frac{1}{\theta}}]$$

where $\theta \geq 1, 0 \leq \delta \geq 1$.

3 Inference Functions for Margins (IFM)

Let $\boldsymbol{\Psi}$ be a vector of functions with the same dimension as parameter vector $\boldsymbol{\alpha}' = (\boldsymbol{\delta}_1', ..., \boldsymbol{\delta}_d', \boldsymbol{\theta}')$. Consider a sample of size T with observed random vectors $\boldsymbol{y_1}, \ldots, \boldsymbol{y_T}$. The vector of *inference functions* is:

$$\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{y}) = \begin{pmatrix} \nabla_{\boldsymbol{\delta_1}} \ell_1(\boldsymbol{\delta_1}; \boldsymbol{y_1}) \\ \vdots \\ \nabla_{\boldsymbol{\delta_d}} \ell_d(\boldsymbol{\delta_d}; \boldsymbol{y_d}) \\ \nabla_{\boldsymbol{\theta}} \ell_c(\boldsymbol{\theta}; \boldsymbol{\delta_1}, ..., \boldsymbol{\delta_d}, \boldsymbol{y}) \end{pmatrix} \tag{5}$$

where $\ell_c(\boldsymbol{\theta}; \boldsymbol{\delta_1}, ..., \boldsymbol{\delta_d}, \boldsymbol{y})$ is the copula log-likelihood, and $\ell_i(\boldsymbol{\delta_i}; \boldsymbol{y_i}) = \sum_{t=1}^{T} \log f(\boldsymbol{\delta_i}; y_{it})$ is the log-likelihood for the univariate margin $i = 1, \ldots, d$. Then, under regularity conditions of score equations in asymptotic maximum likelihood theory, the IFM estimator, $\tilde{\boldsymbol{\alpha}}$, is the root of (5). More specifically, this approach consists of:

(1) implementing d separate optimizations of the marginal log-likelihoods $\ell_i(\boldsymbol{\delta_i}; \boldsymbol{y_i})$ in

the first stage to obtain $\tilde{\boldsymbol{\delta}}_{\mathbf{1}}, ..., \tilde{\boldsymbol{\delta}}_{\boldsymbol{d}}$.

(2) optimizing the copula log-likelihood $\ell_c(\boldsymbol{\theta}; \tilde{\boldsymbol{\delta}}_{\mathbf{1}}, ..., \tilde{\boldsymbol{\delta}}_{\boldsymbol{d}}, \boldsymbol{y})$ over $\boldsymbol{\theta}$ in the second stage to get $\tilde{\boldsymbol{\theta}}$.

Chapter 1

Modelling dependence dynamics between S&P500 and FTSE100 using copula with regime switching specification

1.1 Introduction

Quantifying dependence between random variables is an important area of research in financial econometrics. A conventional measure of association, namely linear correlation coefficient, only measures linear dependence. It is an appropriate measure of dependence only when random variables are represented by elliptical distributions. If this is not the case, then the linear correlation coefficient may provide an inaccurate measure of the dependence structure, and may lead to misleading conclusions (Aas, 2004). Copula functions are powerful tools to capture non-linear dependence between random variables and, therefore, they have become increasingly popular in financial applications. There is growing body of literature that provides evidence of deviations from multivariate normality and asymmetric dependence of equity returns (Longin and Solnik, 2001; Ang and Chen, 2002; Patton, 2006; Ang and Bekaert, 2002).

One of the main factors that exacerbated the financial crisis of 2007 was that banks and financial institutions had built up excessive on- and off-balance sheet leverage (BIS, 2010). This coupled with low capital level of insufficient quality and inadequate liquidity buffers

resulted in the banking system not being able to absorb the ensuing systemic trading and credit losses. Consequently, the loss of confidence in the solvency of financial institutions resulted in a substantial contraction of credit availability and liquidity. The crisis has also spread globally affecting financial markets in other countries. This would question the benefits from international diversification of financial assets.

The financial crisis of 2007 has prompted the Basel Committee to revise and improve the framework of capital requirements. Capital requirement is the amount of capital that a financial institution is required to hold. It is determined as a percentage of a financial institution's asset value, where each asset's value is weighted according to its riskiness.

In this paper we employ Silva Filho et al. (2012) specification to study time-varying dependence structure between S&P500 and FTSE100 indices. This approach identifies two regimes: one of high dependence and one of low dependence. In the low dependence regime we find evidence of asymmetric dependence, in the sense that there is greater dependence in the lower tail than in the upper tail [1]. However, in the high dependence regime we only find a mild asymmetry in the dependence structure. This asymmetry is characterised by an increase in the strength of dependence in both tails, with a relatively greater increase in the upper tail. This is a somewhat surprising result because empirical evidence from various studies on the dependence structure across international financial markets would suggest greater asymmetry during market downturn.

This chapter provides additional evidence on the time-varying dependence structure between the US and the UK financial markets. Second, this chapter considers alternative bootstrap procedure to block bootstrapping.

The paper is structured as follows. In section 1.2 we review the existing literature on copula models with regime switching. Section 1.3 outlines the time varying dependence structure modelling and Section 1.4 presents methodology on marginal distributions. In section 1.5 we discuss estimation procedure. Section 1.6 presents empirical results and concluding remarks are presented in section 1.7.

[1] In this paper we use interchangeably the definition of the asymmetric dependence to mean both that the strength of the dependence may be different across various points in time and, that the strength of the dependence may be different for negative and positive returns.

1.2 Literature review

The combination of copulas and regime switching theory is a relatively recent practice. Jondeau and Rockinger (2006) use regime switching Gaussian and Student-t copula to model the dependence between indices. Rodriguez (2007) models the dependence structure with switching-parameter copulas among daily returns from five East Asian stock indices during the financial crisis, and daily returns from four Latin American stock indices. The results of his study provide evidence that the dependence among stock indices changes during periods of crisis. Garcia and Tsafack (2011) apply regime-switching model to analyse the dependence structure in the international equity and bond markets. The model is based on two regimes: one regime is normal with symmetric dependence and another regime is characterised by asymmetric dependence structure. Doman (2008) analyses the conditional dependence structure between daily returns on the indices listed on the Warsaw Stock Exchange using a Markov-switching copula model. Their results provide additional evidence that the dependence of returns is much stronger during bear markets than in bull markets. Chollete et al. (2009) applies a regime-switching copula to model the dependence structure among stock indices of the G5 and Latin American countries. The analysis extends to higher dimensions and introduces a canonical vine copula. Long (2007) adds regime-switching to the GARCH-copula in order to describe the dependence structure between S&P500 and Nasdaq indices. His estimation results show that regime-switching model outperforms the model without the regime-switching through assessment of log-likelihood, AIC and BIC. Kenourgios et al. (2011) analyse financial contagion using a multivariate regime-switching copula model. The study focuses on four emerging equity markets (Brazil, Russia, India, China) and two developed markets, namely US and UK. In this model, the Gaussian copula is used for the joint distribution, and a GJR-GARCH-MA-t specification for the marginal distributions.

1.3 Specification of time variation

Regime-switching models have been extensively applied in finance since they were introduced in econometrics by Hamilton (1989). Following his approach, it is assumed that a 2-dimensional time series vector $\mathbf{x}_t = (x_{1t}, x_{2t})$, $t = 1, ..., T$, depends on an unobserved binary state variable which indicates the economy's current regime. Furthermore, it is assumed that the latent state variable evolves as a first-order Markov Chain with transition probability defined as follows:

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix} \tag{1.6}$$

where p_{kl} represents the probability of moving to state l at time $t+1$ from state k at time t. In this study we use three copulas presented in Preliminaries section 2. The dependence parameter of each copula is allowed to vary over time according to a specification proposed by Patton (2006), which is extended by allowing the intercept term to switch according to a latent state variable:

$$\theta_{t,S_t} = \Lambda(\omega_c^{S_t} + \beta_c \theta_{t-1,S_{t-1}} + \psi_t) \tag{1.7}$$

where S_t is the latent binary state variable that follows a Markov Chain, $\Lambda(\cdot)$ is the logistic transformation function that constrains the dependence parameter in the appropriate interval, ψ corresponds to a forcing variable. For SJC copula it is defined as the mean absolute differences between the u_1 and u_2[2]:

$$\psi_t = \alpha_c \frac{1}{10} \sum_{j=1}^{10} |u_{1,t-j} - u_{2,t-j}| \tag{1.8}$$

For the Gaussian copula it is given as:

$$\psi_t = \alpha_c \frac{1}{10} \sum_{j=1}^{10} \Phi^{-1}(u_{1,t-j}) \cdot \Phi^{-1}(u_{2,t-j}) \tag{1.9}$$

[2] $U_1 \equiv F_1(X_1)$ and $U_2 \equiv F_2(X_2)$ are the probability integral transforms as in (2) and (3).

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal.

and for Student copula is defined as follows:

$$\psi_t = \alpha_c \frac{1}{10} \sum_{j=1}^{10} t_\nu^{-1}(u_{1,t-j}) \cdot t_\nu^{-1}(u_{2,t-j}) \tag{1.10}$$

where ν is the degree-of-freedom parameter, and $t_\nu^{-1}(\cdot)$ is the inverse of the standard Student-t cumulative distribution function.

1.4 Marginal models

Given a 2-dimensional time series vector $\mathbf{x}_t = (x_{1t}, x_{2t})$, $t = 1, ..., T$, the copula-TGARCH model can be represented as follows:

$$H(\mathbf{x}|\boldsymbol{\mu}, \mathbf{h_t}) = C_{\theta_t}(F_1(x_{1t}|\mu_1, h_{1t}), F_2(x_{2t}|\mu_2, h_{2t})|\boldsymbol{\mu}, \boldsymbol{h}; \theta_{t,S_t}) \tag{1.11}$$

where C_{θ_t} is the time-varying copula with parameter θ_{t,S_t}, and $F_i(x_{it}|\mu_i, h_{it}) \quad i = 1, 2$, are the marginal distributions specified as a univariate TGARCH processes.

The evolution of a TGARCH(P,Q) process can be described by the following model

$$x_{it} = \mu_i + u_{it} \quad \text{where} \quad u_{it} = h_{it}^{\frac{1}{2}} \epsilon_{it} \tag{1.12}$$

$$h_{it} = \omega_i + \sum_{p=1}^{P} \alpha_{ip} u_{it-p}^2 + \sum_{o=1}^{O} \gamma_{io} u_{it-o}^2 I_{[u_{it-o}<0]} + \sum_{q=1}^{Q} \beta_{iq} h_{it-q} \tag{1.13}$$

where h_{it} is the conditional variance given past information, and ϵ_{it} are $i.i.d.$ random variables, ω_i, β_i, $\alpha_i > 0$ assures that the conditional variance h_{it} is positive and $\alpha_i + \beta_i < 1$ ensures covariance stationarity. It is also assumed that the ϵ_{it} has a skewed-t distribution, with its density given by:

$$f(\epsilon_{it}|\upsilon, \lambda) = \begin{cases} bc\left(1 + \frac{1}{\upsilon - 2}\left(\frac{b\epsilon_{it} + a}{1 - \lambda}\right)^2\right)^{\frac{-(\upsilon+1)}{2}} & \epsilon_{it} < \frac{-a}{b} \\ bc\left(1 + \frac{1}{\upsilon - 2}\left(\frac{b\epsilon_{it} + a}{1 + \lambda}\right)^2\right)^{\frac{-(\upsilon+1)}{2}} & \epsilon_{it} \geq \frac{-a}{b} \end{cases}$$

where constants a, b and c are defined as:

$$a = 4\lambda c\left(\frac{\upsilon-2}{\upsilon-1}\right), \quad b^2 = 1+3\lambda^2-a^2, \quad c = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\pi(\upsilon-2)\Gamma\left(\frac{\upsilon}{2}\right)}}$$

with υ corresponding to the number of degrees of freedom, and λ representing the degree of asymmetry. When λ is negative, we have a left-skewed density, meaning that there is higher probability of observing large negative returns than large positive returns. We follow Chollete et al. (2009) and Silva Filho et al. (2012) in that we model the marginal distributions independent of the regime.

1.5 Estimation

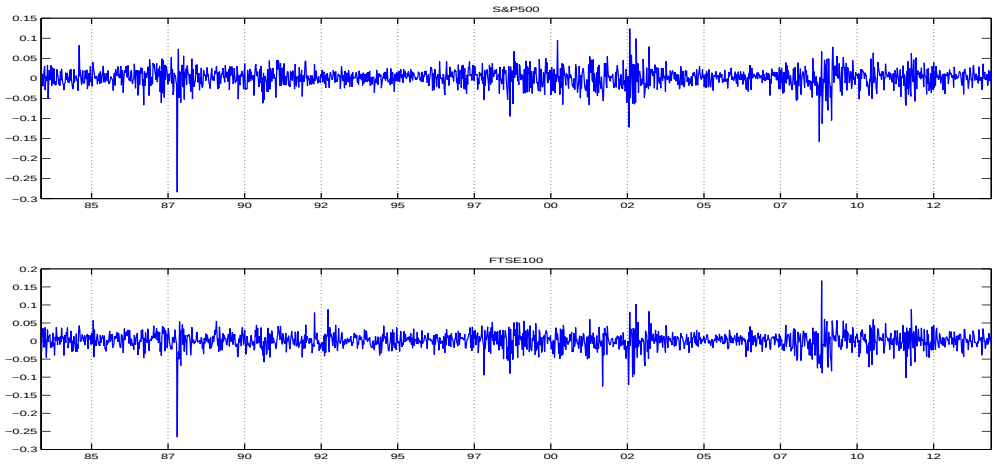
The log-likelihood function has the following form:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{t=1}^{T}\log\left(c_{\theta_t}(F_1(x_{1t}|\mu_1,h_{1t}),F_2(x_{2t}|\mu_2,h_{2t})|\theta_{t,S_t})\times\prod_{i=1}^{2}f_{it}(x_{it}|\mu_{it},h_{it})\right) \tag{1.14}$$

Since the log-likelihood function in (1.14) is a separable function, the maximum likelihood estimation procedure can be performed in two steps. This method known as the Inference Functions for Margins (IFM) was proposed by Joe and Xu (1996). In the first step of this procedure, the parameters of the univariate marginal distributions are estimated. In the second step, these estimates are used to estimate the copula parameters. The log-likelihood in (1.14) can be rewritten as:

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{x}) &= \sum_{t=1}^{T}\log\left(c_{\theta_t}\left(F_1(x_{1t}|\mu_1,h_{1t}),F_2(x_{2t}|\mu_2,h_{2t})|\boldsymbol{\mu},\boldsymbol{h};\theta_{t,S_t}\right)\times\prod_{i=1}^{2}f_{it}(x_{it}|\mu_{it},h_{it})\right) \\ &= \sum_{t=1}^{T}\log c_{\theta_t}\left(F_1(x_{1t}|\mu_1,h_{1t}),F_2(x_{2t}|\mu_2,h_{2t})|\boldsymbol{\mu},\boldsymbol{h};\theta_{t,S_t}\right)+ \\ &\quad +\sum_{t=1}^{T}\log f_{1t}(x_{1t}|\mu_1,h_{1t})+\sum_{t=1}^{T}\log f_{2t}(x_{2t}|\mu_2,h_{2t}) \end{aligned} \tag{1.15}$$

Figure 1.1: Weekly Logarithmic Returns



The copula part can be rewritten as:

$$\sum_{t=1}^{T} \log c_{\theta_t}\left(u_{1t}, u_{2t} | \boldsymbol{\mu}, \boldsymbol{h}; \theta_{t,S_t}\right) = \sum_{t=1}^{T} \log \left(\sum_{S_t=1}^{2} c_{\theta_t}(u_{1t}, u_{2t} | \boldsymbol{\mu}, \boldsymbol{h}; S_t, w_{t-1}) Pr(S_t | w_{t-1}) \right) \quad (1.16)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\boldsymbol{h} = (h_1, h_2)'$. Due to the state variable S_t being unobservable, in order to evaluate the log-likelihood we need to calculate the weights $Pr(S_t | w_{t-1})$ for $S_t = 1$ and $S_t = 2$. The procedure is to apply filter as in Hamilton (1994) and Kim and Nelson (1999), which yields the following algorithm that should be iterated through $t = 1, ..., T$:

Step 1: *Prediction of* S_t

$$Pr(S_t = l | w_{t-1}) = \sum_{k=1}^{2} p_{kl} Pr(S_{t-1} = k | w_{t-1})$$

for $l = 1, 2$ and $p_{kl} = Pr(S_t = l | S_{t-1} = k)$ is the transition probability between the states k and l as introduced in (1.6).

Step 2: *Filtering of* S_t

$$Pr(S_t = l | w_t) = \frac{c_t(u_1, u_2 | S_t = l, w_{t-1}) Pr(S_t = l | w_{t-1})}{\sum_{k=1}^{2} c_t(u_1, u_2 | S_t = k, w_{t-1}) Pr(S_t = k | w_{t-1})}$$

Figure 1.2: Sample ACF of Returns

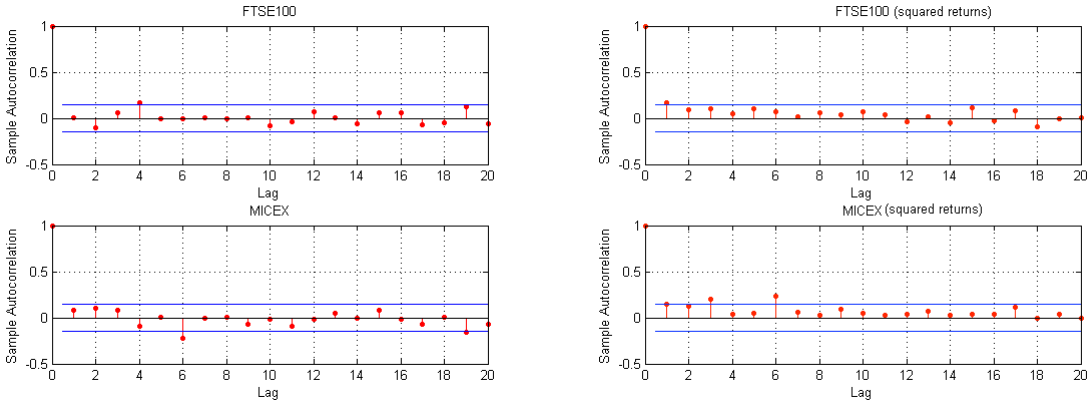
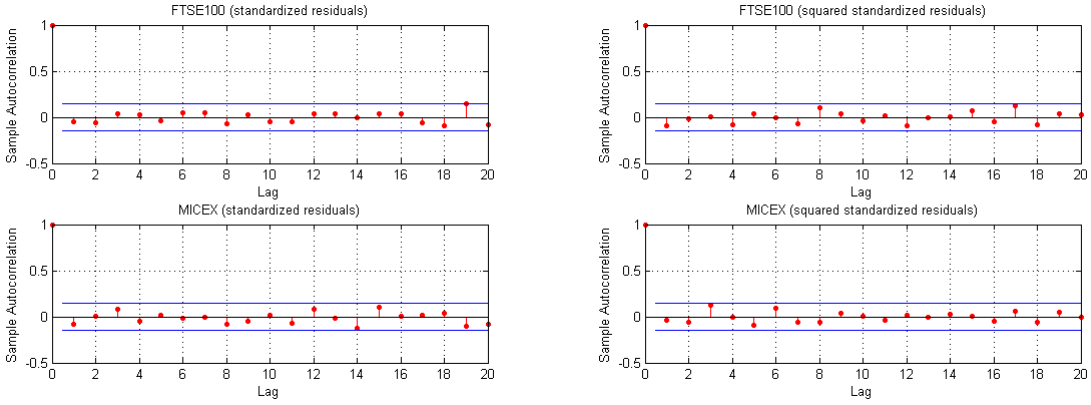


Figure 1.3: Sample ACF of Standardized Residuals



where $w_t = (w_{t-1}, u_{1t}, u_{2t})'$. At $t = 1$ the filter is initialized using stationary probabilities of S_t:

$$\pi_1 = Pr(S_0 = 1|w_0) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \tag{1.17}$$

$$\pi_2 = Pr(S_0 = 2|w_0) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \tag{1.18}$$

Consequently, using this filter the probability distribution of S_t is obtained given the information set up to time t. However, it would be useful to know the distribution of S_t given the full sample information set, that is using all T observations. Therefore, the smoothed probabilities for S_t, $Pr(S_t = l|w_T) = \sum_{k=1}^{2} Pr(S_t = l, S_T = k|w_T)$ can be calculated recursively from the filtered probabilities. This smoothing process is carried out in the following way as outlined in Hamilton (1994) and Kim and Nelson (1999):

Table 1.1: Descriptive Statistics

	S&P500	**FTSE100**
Mean	0.001580	0.001291
Median	0.003399	0.002214
Maximum	0.123746	0.166889
Minimum	-0.283705	-0.265825
Std. Deviation	0.023637	0.023848
Skewness	-1.504287	-1.100278
Kurtosis	18.64870	15.66106
Jarque-Bera	17330.95	11271.14
	(0.000000)	(0.000000)

Note: Jarque-Bera corresponds to Jarque-Bera test statistics with p-values in parentheses.

1. Filtered probabilities $Pr(S_t = l|w_t)$ are obtained for $l = 1, 2$ and $t, .., T$.

2. The smoothing algorithm is initialized at time $t = T$ with $Pr(S_T = l|w_T)$

3. The smoothed probability distribution for each $t = T - 1, ..., 1$ is obtained by:

$$Pr(S_t = l|w_T) \quad = \quad Pr(S_t = l|w_t) \sum_{k=1}^{2} \frac{p_{lk} Pr(S_{t+1} = k|w_T)}{\sum_{j=1}^{2} p_{jk} Pr(S_t = j|w_t)} \tag{1.19}$$

   where $p_{lk} = Pr(S_{t+1} = k|S_t = l)$ are the transition probabilities between the states l and k.

At $t = 0$ the smoothing algorithm calculates $Pr(S_0 = l|w_T)$ which can be used in the filtering process as the initial value. Subsequently, equation (1.16) can be maximised numerically with respect to the model parameters following.

1.5.1 Estimation of Standard Errors

Once the estimates of the copula parameters are obtained, it would be of interest to compute standard errors of those estimates, which can subsequently provide an idea of the precision with which these parameters have been estimated. In what follows we discuss one possible procedure to compute standard errors of the estimates.

Standardized residuals bootstrap (SRB)

The following algorithm is used to calculate standard errors using SRB procedure:

Table 1.2: Marginal Distributions Estimation Results

	S&P500	FTSE100
ω_i	0.0009	0.0001
	(0.000001)	(0.000001)
α_i	0.0340	0.1170
	(0.0002)	(0.0006)
γ_i	0.1343	–
	(0.0011)	–
β_i	0.8750	0.8287
	(0.0007)	(0.0011)
υ_i	9.0788	6.7236
	(5.9554)	(1.8111)
λ_i	- 0.2712	– 0.1788
	(0.0012)	(0.0012)
$\log L$	4089.0	3981.3
$Q(10)$	0.3016	0.4923
$Q^2(20)$	1.0000	1.0000
KS	0.9288	0.9841
$Berk$	0.9951	0.6802

Note: In round brackets are the standard errors. Q and Q^2: Ljung-Box Q statistic for autocorrelation and squared autocorrelation in the residual terms, respectively. Last two rows show goodness-of-fit tests for the probability integral transform of the margins. KS: Kolmogorov - Smirnov test p-value, and $Berk$: Berkowitz test p-value for uniformity.

1. Sample a pair $\boldsymbol{e}_t^* = (e_{1t}^*, e_{2t}^*)$ with replacement from the original standardized *i.i.d.* residuals $\{\boldsymbol{e}_t \ldots, \boldsymbol{e}_T\}$ obtained from marginal models in (1.12) and (1.13), and produce bootstrapped sample $\boldsymbol{e}^b = (\boldsymbol{e}_1^* \ldots, \boldsymbol{e}_T^*)'$.

2. Transform each bootstrapped time series data $\boldsymbol{e}_i^* = (e_{i1}^* \ldots, e_{iT}^*)'$ for $i = 1, 2$ using skewed-t cumulative distribution function to produce $\boldsymbol{u}_i^* = (u_{i1}^* \ldots, u_{iT}^*)'$ for $i = 1, 2$.

3. Estimate parameter vector $\boldsymbol{\theta}_b^*$ of the regime-switching copula model using $\boldsymbol{u}^b = (\boldsymbol{u}_1^* \ldots, \boldsymbol{u}_T^*)'$ where $\boldsymbol{u}_t^* = (u_{1t}^*, u_{2t}^*)$.

4. Repeat steps (1) to (3) B times.

5. Calculate the standard errors for the parameters using the covariance matrix $(B - 1)^{-1} \sum_{b=1}^{B} (\boldsymbol{\theta}_b^* - \overline{\boldsymbol{\theta}^*})(\boldsymbol{\theta}_b^* - \overline{\boldsymbol{\theta}^*})'$, where $\overline{\boldsymbol{\theta}^*}$ denotes the mean of the estimates across B bootstrapped samples.

In order to maintain the cross-sectional dependence of the data, the above procedure is carried out on entire rows of the standardized residuals.

Table 1.3: Copula estimation results

Coefficient	**Gaussian**	**Student**	Coefficient	**Symmetrised Joe-Clayton**
ω_c^1	3.8150	3.8945	$\omega_{c,U}^1$	0.9557
	(0.9986)	(1.9123)		(2.8364)
ω_c^2	1.9227	2.2338	$\omega_{c,U}^2$	-0.9758
	(0.3160)	(0.0983)		(1.0582)
β_c	-2.1964	-2.3708	$\beta_{c,U}$	3.8034
	(1.2445)	(0.9877)		(1.9964)
α_c	0.2737	0.0496	$\alpha_{c,U}$	-0.9094
	(0.2414)	(0.0800)		(1.3791)
υ^1	–	7.2865	$\omega_{c,L}^1$	-3.1403
	–	(2.0907)		(2.2521)
υ^2	–	7.1505	$\omega_{c,L}^2$	2.3388
	–	(1.5282)		(1.7904)
	–	–	$\beta_{c,L}$	-10.2920
	–	–		(6.8713)
	–	–	$\alpha_{c,L}$	-1.3621
	–	–		(2.2746)
p_{11}	0.9878	0.9993	p_{11}	0.9993
	(0.4145)	(0.4268)		(0.0748)
p_{22}	0.9853	0.9992	p_{22}	0.9993
	(0.4141)	(0.2322)		(0.0748)
log L	**454.8**	**468.8**		**460.8**

Note: Standard errors in round brackets are estimated using standardized residuals bootstrap procedure (SRB) as outlined in 1.5.1. The transition probability parameters p_{11} and p_{22} are defined in (1.6).

1.6 Empirical application

1.6.1 Data

The monthly data used for this study comprises S&P500 and FTSE100 indices for the period May 10, 1983 to May 20, 2014, which gives us 1620 observations. The source of the data is Datastream. The returns are logarithmic monthly returns. Table 1.1 presents summary statistics of the data. Both means are much smaller relative to their standard deviation. The S&P500 returns exhibit substantial skewness and high kurtosis, whilst FTSE100 exhibits a relatively lower skewness and kurtosis. The Jarque-Bera test of normality of each return series strongly rejects the null, indicative of non-normality of both series.

1.6.2 Empirical Results

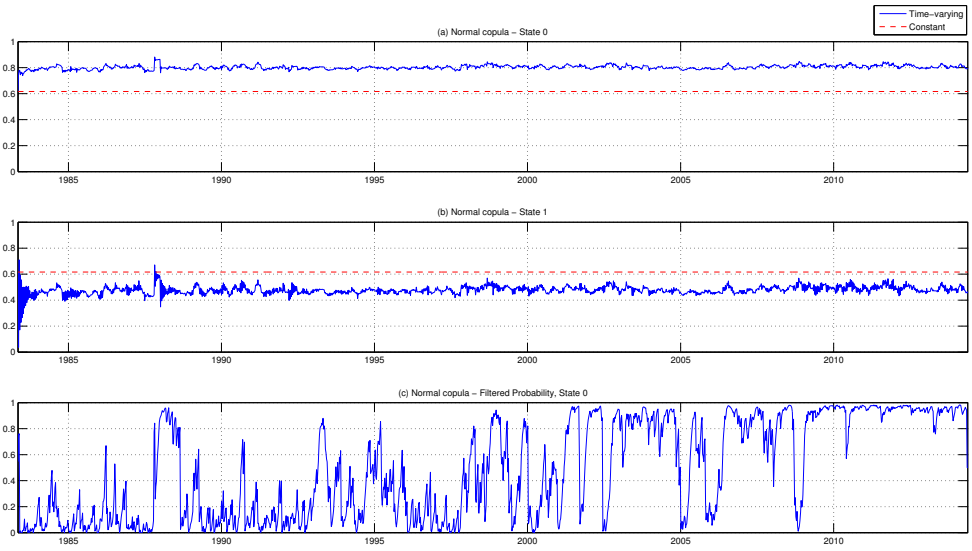
Table 1.2 shows estimation results for marginal models. The asymmetry coefficient is negative for both series. This indicates that both series are negatively skewed. That is, large negative returns are more likely than large positive returns, which is consistent with Figure 1.1. The degrees of freedom parameters for S&P500 and FTSE100 are approximately 9 and 6 respectively. Low estimates indicate presence of heavy tails in both distributions. In addition high estimates of the β coefficients for both series suggests high volatility persistence. The reported p-values for the Ljung-Box Q statistic and sample ACF plots in Figure 1.2 and 1.3 indicate both no autocorrelation and no squared autocorrelation in the standardised residual terms. The results for Kolmogorov-Smirnov (KS) and Berkowitz (Berk) tests for uniformity of the transformed standardised residuals are also reported. High p-values suggest no evidence against the null of Uniform(0,1) distribution of the transformed residuals.

Table 1.3 presents estimation results for the three copulas with regime-switching specification. High values for the transition probability parameters p_{11} and p_{22} indicate high persistence of both regimes. The magnitude of the intercept in the dynamics of the dependence parameter changes substantially between the two regimes for both Gaussian and Student copula. For the Symmetrised Joe-Clayton copula, not only the magnitude of the difference is considerable, but also the sign changes. Based on the AIC and BIC criteria, the Student-t copula model provides a better fit to the data, which suggests presence of heavy tails in the joint distribution of asset returns. Hence, the Student-t regime switching copula model is the preferred model.

Figures 1.4, 1.5 and 1.6 show the evolution of the dependence parameters in two states, which is displayed in Panels (a) and (b). The probability of being in high dependence state is displayed in Panel (c).

Figure 1.4 displays the evolution of the dependence parameters for the Gaussian copula. In the Gaussian copula, the dependence parameter is the linear correlation coefficient. In the low dependence regime, the parameter oscillates between values of 0.4 and 0.5. In the high dependence regime, it increases noticeably and remains between values of 0.75 and

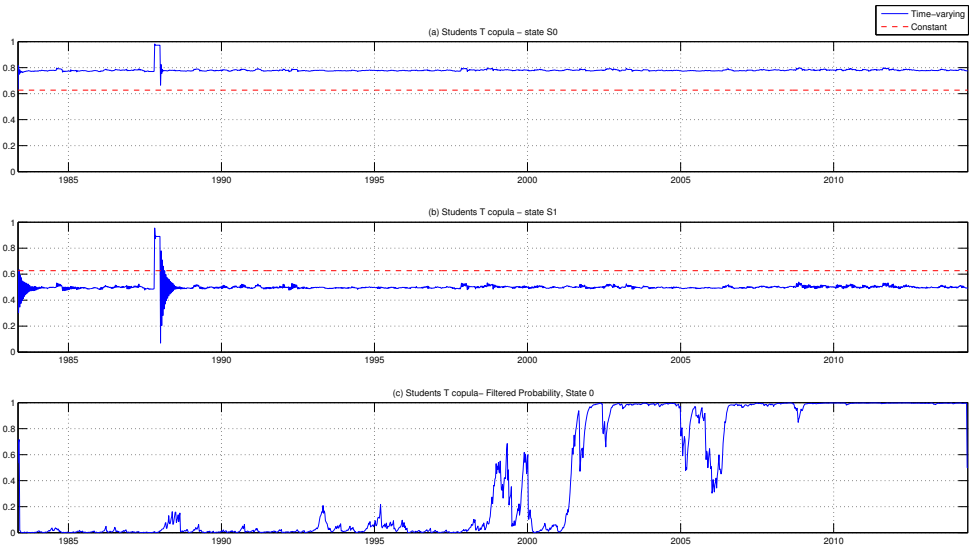
Figure 1.4: Normal Copula



0.85. In Panel (c) it can be seen that there are three visible shifts into high dependence regime before 2000. The first jump occurs starting from mid 1987 until mid 1988. It is interesting to note that this period coincides with the October 19, 1987 stock market crash which is referred to as "Black Monday". The market crash of 1987 was characterised by a swift and severe market decline, with the S&P500 and FTSE100 falling approximately 20 and 11 per cent respectively just in one day. Second jump occurs in 1992 and lasts for less than a year. During this period the British Conservative government suspended its membership of the European Exchange Rate Mechanism (ERM) by withdrawing the pound sterling from the monetary system. This day is often referred to as "Black Wednesday". The third notable jump takes place in 1998, which is preceded by a series of frequent jumps to and from the high dependence regime. This coincides with a sequence of world financial and economic events in the late 1990s: the 1997 Asian and 1998 Russian financial crises. The Asian financial crisis was initiated by rounds of currency depreciation which exposed serious problems in the financial sector. This would, through multiple channels, adversely affect the United States and the United Kingdom. The Russian financial crisis resulted in the Russian government devaluing its currency, defaulting on all domestic debt and declaring moratorium on its foreign debt.

Figure 1.5 presents results for the Student copula. Panel (a) shows that the parameter

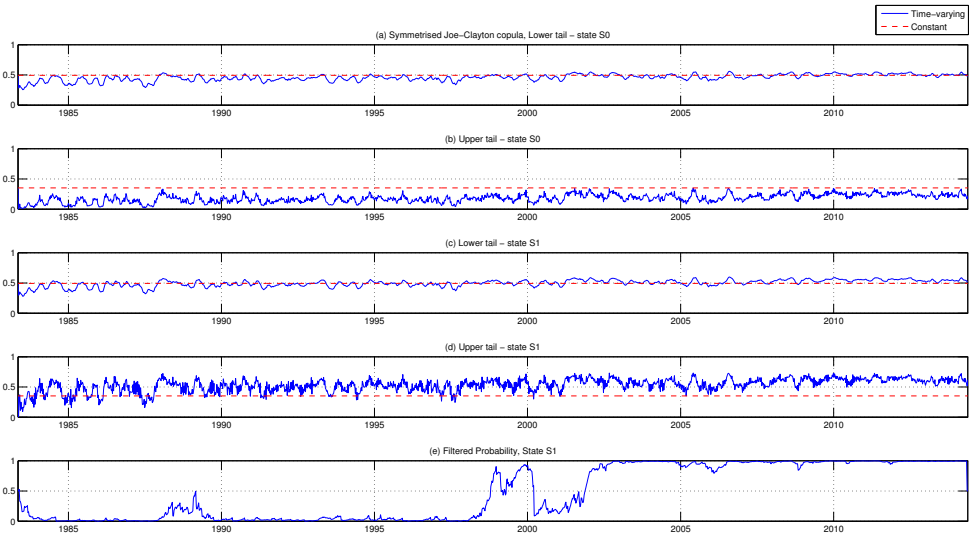
Figure 1.5: Student Copula



in the low dependence regimes oscillates between 0.75 and 0.8, with a large jump in the second half of 1987. In the high dependence regime this parameters varies around 0.45 and 0.55 and experiences a similar jump in the second half of 1987, followed by a large outbursts of volatility that gradually settles down by the third quarter of 1988. Panel (c) displays the filtered probability of being in high dependence regime. Before the 2000 it can be seen that the probability of being in high dependence regime remains close to zero. However, there can be observed jumps into high dependence regime in 1998. After 2001 there is a shift into the high dependence regime which persist until 2014 with a brief return into the low dependence regime at the beginning of 2006. The probability of being in a high dependence regime coincides most of the time with that identified in the Gaussian copula case after year 2000.

Figure 1.6 presents results for the Symmetrised Joe-Clayton copula, which captures dependence both in the upper and lower tail, without imposing them to be the same. In the low dependence regime, the lower tail dependence parameter evolves around a constant parameter value of 0.5. In the high dependence regime, the evolution of this parameter seems not to change substantially. The upper tail dependence parameter, on the other hand, remains below the constant parameter value of 0.36 in the low dependence regime, but increases notably and oscillates around 0.5 most of the time in the high dependence

Figure 1.6: Symmetrised Joe-Clayton Copula



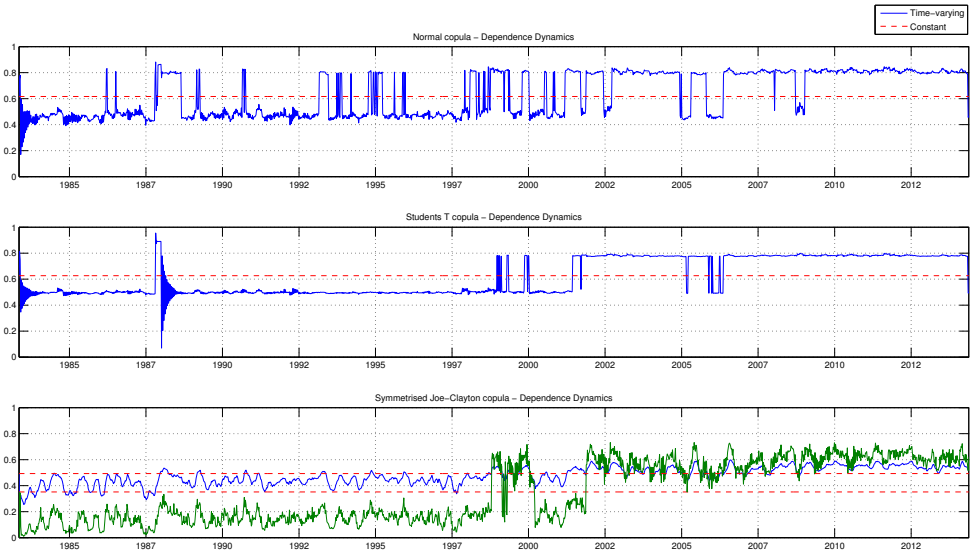
regime. The probability of being in a high dependence regime follows closely the probability of being in a high dependence regime identified in the Student-t copula case.

Figure 1.7 displays the evolution of the dependence parameters of all three copulas, which is constructed as follows: if the probability of being in a high dependence regime exceeds the value of 0.5, the evolution of the parameters from the high dependence regime is plotted, and if it is equal or below the value of 0.5, then the evolution from the low dependence regime is plotted. Panel displaying the dependence dynamics for the Symmetrised Joe-Clayton copula suggests that during the low dependence regime, the dependence is stronger in the lower tail than in the upper tail. However, in the high dependence regime, the dependence in the upper tail increases and remains close to the dependence in the lower tail. This result would seem to indicate that the high dependence regime is characterised by a low asymmetry in the upper and lower tails compared to the low dependence regime. The superimposed dashed red line represents an estimate of the corresponding static copula parameters.

1.7 Conclusion

In this study we investigated time-varying dependence structure between S&P500 and FTSE100. In order to model marginal distributions AR skew-t-GARCH and AR skew-

Figure 1.7: Dependence dynamics



t-TGARCH models were employed. To model the dependence structure three copulas were estimated: the Gaussian copula, the Student-t copula, and the Symmetrized Joe-Clayton copula. The Student-t regime switching copula model provides a better fit to the data, which suggests existence of tail dependence, and presence of heavy tails in the joint distribution of asset returns. This indicates the possibility of joint extreme losses happening together. Time variation was modelled by allowing the parameter in each copula to evolve according to Patton (2006) framework. Regime switching specification was introduced by allowing the intercept in the evolution process of the parameter to vary according to first-order Markov chain. This modelling approach identified two regimes over the sample period: one of low dependence and one of high dependence. In the low dependence regime we find a relatively higher dependence in the lower tail than in the upper tail, indicative of the dependence asymmetry. In the high dependence regime we detect an increase in the dependence in both tails, with a relatively stronger increase in the upper tail.

Chapter 2

Modelling dependence structure between stock indices using a second-order regime-switching vine copula

2.1 Introduction

It is now widely accepted that the hypothesis of multivariate normality is not usually supported by the data for financial time series. Vine copulas offer an alternative to elliptical multivariate distributions that can be constructed in a tractable and flexible way. Joe (1996) showed that multivariate distribution functions could be represented using pair-copula construction (PCC). This class of structure is based on the decomposition of the multivariate density into a product of marginal densities and bivariate copulae, which can be from any family. This allows greater flexibility as several families may be mixed in one pair-copula construction. However, for high-dimensional distributions there is a significant number of different possible decompositions. To organize these numerous decompositions, Bedford and Cooke (2001) introduced graphical models that are composed of a sequence of trees called regular vines, which share some similarities to Bayesian networks. Subsequently Aas et al. (2009) considered two classes of regular vine copulas called canonical vine (C-vine) and the drawable vine (D-vine), and set them in an inferential context. Since financial markets undergo episodes of rapid growth and dramatic decline of stock prices, adequate models able to capture variations in dependence structure are required. The objective of this chapter is to extend first order Markov switching R-vine model to a higher order

regime-switching model. In addition, it is of interest to investigate any advantages of regime switching vine copula relative to static vine copula model.

The structure of this chapter is as follows. Section 2.2 presents literature review and in the Section 2.3 the required methodological background on vine copulas is presented. The estimation results for the regime-switching vine copula models are presented in Section 2.6. Section 2.7 provides concluding remarks.

2.2 Literature review

The literature combining copulas and regime switching theory experienced growth in the last decade. However, most of studies focus on the application of Markov switching theory to bivariate copula families and very few to higher dimensions. This is associated with the difficulties of building multivariate distributions.

In the bivariate context, Rodriguez (2007) models the dependence structure with switching-parameter copulas among daily returns from five East Asian stock indices during the financial crisis, and daily returns from four Latin American stock indices. The results of his study provide evidence that the dependence among stock indices changes during periods of crisis. Garcia and Tsafack (2011) applies regime-switching model to analyse the dependence structure in the international equity and bond markets. The model is based on two regimes: one regime is normal with symmetric dependence and another regime is characterised by asymmetric dependence structure. Long (2007) adds regime-switching to the GARCH-copula in order to describe the dependence structure between SP500 and Nasdaq indices. His estimation results show that regime-switching model outperforms the model without the regime-switching through assessment of log-likelihood, AIC and BIC. Kenourgios et al. (2011) analyse financial contagion using a multivariate regime-switching copula model. The study focuses on four emerging equity markets (Brazil, Russia, India, China) and two developed markets, namely US and UK. Doman (2008) analyses the conditional dependence structure between daily returns on the indices listed on the Warsaw Stock Exchange using a Markov-switching copula model. Their results provide additional evidence that the dependence of returns is much stronger during bear markets than in bull

markets.

In the multivariate context, the first use of Markov switching theory combined with C-vine copula was by Chollete et al. (2009). In their study the authors modelled time varying dependence structure among stock indices of the G5 and Latin American countries. Stöber and Czado (2014) extended initial approach by Chollete et al. (2009) to a general R-vine model. In that study, the authors investigate the dependence structure of 9 exchange rates against US Dollar, the dependence of daily log returns for 5 stock indices of the eurozone and the dependence structure of daily log returns of individual stocks selected from a German stock market. They find that regime switches are present in all kinds of financial time series data sets and, therefore, Markov switching models could offer a more accurate description of time varying dependence structure during different regimes. In this chapter we extend these modelling frameworks to a second-order regime-switching vine copula model.

2.3 Vine Copula Theory

In this section we present the theory of regular vine copula that is used to study the dependence structure amongst three financial stock indices.

2.3.1 Regular Vine Copulas

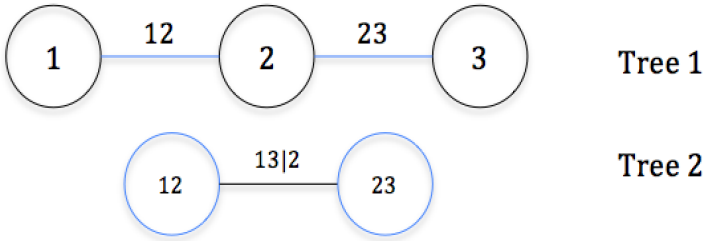
The approach used to organise the pair-copula construction (PCC) is known as regular vine tree sequence. A tree is an undirected graph in which any two nodes are connected by exactly one edge. In other words a tree is a connected, undirected, acyclic graph. In this construction, nodes denote standard uniform random variables, and edges represent the appropriate pair-copula densities for the pair of uniform random variables.

Definition 1.1 A set of linked trees $\mathcal{V} = \{T_1, ..., T_{d-1}\}$ is a regular vine on n elements if:

1. T_1 is a tree with nodes $N_1 = \{1, ..., d\}$ and a set of edges E_1.

2. T_i is a tree with nodes $N_i = E_{i-1}$ and edge set E_i for $i = 2, ..., d-1$.

Figure 2.8: Graphical representation of a 3-dimensional vine copula



3. If $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are two nodes in N_i connected by an edge, then exactly one element of a equals one element of b for $i = 2, ..., d-1$.

The third condition ensures that the decomposition into bivariate copulas is well defined (Czado et al., 2012).

Regular vine copula

A regular vine on d variables is one in which two edges in tree j are joined by an edge in tree $j+1$ only if these edges share a common node (Kurowicka and Cooke, 2006). The complete union A_e of an edge $e = \{a, b\} \in E_i$ in tree T_i of a regular vine $\mathcal{V}$ is defined as:

$$A_e = \{v \in N_1 : \exists e_m \in E_m, m = 1, ..., i-1, \quad \text{such that} \quad v \in e_1 \in \cdots \in e_{i-1} \in e\}$$

That is, the complete union of an edge e is simply a set of all nodes in tree T_1 that can be reached from that particular edge e. The conditioning set of an edge $e = \{a, b\}$ is defined as $D_e := A_a \cap A_b$ and the conditioned set associated with $e = \{a, b\}$ are defined as $\mathscr{C}_{e,a} = A_a \backslash D_e$ and $\mathscr{C}_{e,b} = A_b \backslash D_e$. It can be shown that the conditioned sets are singletons and, therefore, we can refer to edges by their labels $\{\mathscr{C}_{e,a}, \mathscr{C}_{e,b} | D_e\} := \{i(e), j(e) | D(e)\}$, where $i(e)$ and $j(e)$ are two nodes connected by an edge e. Subsequently, a regular vine copula can be specified, given these sets, by associating a (conditional) bivariate copula with every edge of the regular vine.

Definition 1.2 A regular vine copula $C = ((\mathcal{V}), B(\mathcal{V}), \boldsymbol{\theta}(B(\mathcal{V})))$ in d dimensions is a multivariate distribution function such that for a random vector $\boldsymbol{U} = (U_1, ..., U_d)' \sim C$ with uniform margins:

1. $\mathcal{V}$ is a regular vine on n elements.

2. $B(\mathcal{V}) = \{C_{i(e),j(e)|D(e)} | e \in E_m, m = 1, ..., d-1\}$ is a set of $d(d-1)/2$ copula families that identify the conditional distributions of $U_{i(e)}, U_{j(e)} | \boldsymbol{U}_{D(e)}$.

3. $\boldsymbol{\theta}(B(\mathcal{V})) = \{\boldsymbol{\theta}_{i(e),j(e)|D(e)} | e \in E_m, m = 1, ..., d-1\}$ is the set of parameter vectors associated with the copulas in $B(\mathcal{V})$.

The probability density function $f_{1:d}$ of $\mathbf{x} = (x_1, \ldots, x_d)' \in \mathbb{R}^d$ of a d-dimensional regular vine distribution $F_{1:d}$ is as follows:

$$f_{1:d}(\boldsymbol{x}|\mathcal{V}, B, \boldsymbol{\theta}) = \left[\prod_{m=1}^{d-1} \prod_{e \in E_m} C_{i(e),j(e)|D(e)} \left(F_{i(e)|D(e)}, F_{j(e)|D(e)} | \boldsymbol{\theta}_{i(e),j(e)|D(e)}\right)\right] \times \left[\prod_{k=1}^{d} f_k(x_k)\right] \tag{2.20}$$

where $F_{i(e)|D(e)} := F_{i(e)|D(e)}\left(x_{i(e)}|\boldsymbol{x}_{D(e)}\right)$ and $F_{j(e)|D(e)} := F_{j(e)|D(e)}\left(x_{j(e)}|\boldsymbol{x}_{D(e)}\right)$. These conditional distribution functions are determined as follows:

$$\begin{aligned} F_{i(e)|D(e)}\left(x_{i(e)}|\boldsymbol{x}_{D(e)}\right) &= F_{\mathscr{C}_{e,a}|D_e}\left(x_{\mathscr{C}_{e,a}}|\boldsymbol{x}_{D_e}\right) \\ &= \frac{\partial C_{\mathscr{C}_{a,a_1},\mathscr{C}_{a,a_2}}\left(F_{\mathscr{C}_{a,a_1}|D_a}\left(x_{\mathscr{C}_{a,a_1}}|\boldsymbol{x}_{D_a}\right), F_{\mathscr{C}_{a,a_2}|D_a}\left(x_{\mathscr{C}_{a,a_2}}|\boldsymbol{x}_{D_a}\right)\right)}{\partial F_{\mathscr{C}_{a,a_2}|D_a}\left(x_{\mathscr{C}_{a,a_2}}|\boldsymbol{x}_{D_a}\right)} \end{aligned}$$

The likelihood L of a regular vine copula $C = (\mathcal{V}, B, \boldsymbol{\theta})$ given the observed data $\boldsymbol{x} = (\boldsymbol{x_1}, ..., \boldsymbol{x_N})' \in \mathbb{R}^{N \times d}$ is as follows:

$$L(\mathcal{V}, B, \boldsymbol{\theta}|\boldsymbol{x}) = \prod_{k=1}^{N} f_{1:d}(\boldsymbol{x_k}|\mathcal{V}, B, \boldsymbol{\theta}) \tag{2.21}$$

2.3.2 Decomposition of a three dimensional distribution.

The possible number of different D - and C - vines becomes increasingly large as dimensions d increase. Aas et al. (2009) showed that for dimension d there are $d!/2$ distinct C-vine trees and $d!/2$ distinct D-vine trees. The number of possible trees is even larger for regular vines (Morales Napoles et al., 2010). If returns of particular stock index drives returns of all other stock indices then a C-vine tree structure may be reasonable. In other

cases, it may be reasonable to consider D-vine tree. Therefore, in order to keep the process of model selection among vine specification feasible, we restrict our study to three financial returns series. When $d = 3$, C− and D− vine coincide and this simplifies the decision regarding which vine tree structure to choose from. In fact, there are three possible different vine trees in total.

If three continuous random variables $\boldsymbol{X}_{1:3} = (x_1, x_2, x_3)' \in \mathbb{R}^3$ have joint density $f_{1:3}$ with marginal densities f_1, f_2 and f_3 then we can decompose the joint density by conditioning:

$$f_{1:3}(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1,2}(x_3|x_1, x_2) \tag{2.22}$$

The conditional densities in (2.22) can be decomposed further using Sklar's theorem as follows:

$$\begin{aligned} f_{2|1}(x_2|x_1) &= \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} \qquad (2.23) \\ &= c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \\ f_{3|1,2}(x_3|x_1, x_2) &= \frac{f_{1,3|2}(x_1, x_3|x_2)}{f_{1|2}(x_1|x_2)} \qquad (2.24) \\ &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot \\ &\quad \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot f_3(x_3) \end{aligned}$$

Thus, the three-dimensional multivariate density can be expressed as a product of pair-wise copulas for (X_1, X_2) and for (X_2, X_3), pair-wise conditional copula for (X_1, X_3) given X_2, and individual marginal densities for X_1, X_2 and X_3:

$$\begin{aligned} f_{1:3}(x_1, x_2, x_3) &= c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot \qquad (2.25) \\ &\quad \cdot c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \end{aligned}$$

This decomposition is a particular case of both a C- and a D - vine copula, and is unique up to re-labelling of the variables. Figure 2.8 shows the graphical specification corresponding to the three-dimensional C− and D−vine. It consists of two trees $T_j,\ j = 1, 2$. There are three nodes and two edges in tree T_1, and two nodes and one edge corresponding to a

conditional pair-copula in T_2.

2.4 The models for copula

In this section we will use 10 copulas presented in Preliminaries section 2 as building blocks in our vine specifications. In our candidate set of vine copulas, we consider copula families with different tail behaviour. Clayton and Survival Gumbel copulas have only lower tail dependence, whilst BB6 copula has upper tail dependence only. Gaussian, Frank and BB8 copulas have no tail dependence. Student-t, BB1 and BB7 have upper and lower tail dependence, although Student-t copula imposes symmetric dependence in both tails.

2.4.1 Markov Switching

In order to capture variations in the magnitude of the dependence amongst financial stock indices, we introduce a Markov regime-switching specification into vine copula model. The Markov regime-switching vine (MS-V) copula can be characterized by specifying conditional densities of the transformed uniform margins as follows Stöber and Czado (2014):

$$c(\boldsymbol{u_t}|(\boldsymbol{\mathcal{V}}, \mathbf{B}, \boldsymbol{\theta})_{1,\dots,p}, S_t) \quad = \quad \sum_{k=1}^{p} \mathbb{1}_{\{k\}}(S_t) \cdot c(\boldsymbol{u_t}|(\boldsymbol{\mathcal{V}}, \mathbf{B}, \boldsymbol{\theta})_k) \tag{2.26}$$

Thus, we assume that vector $\boldsymbol{u_t}$ at time t depends on a latent variable that indicates the economy's current regime. In other words, regime determines what the current copula is: once we know what the current regime is, we know the copula for that t. Furthermore, it is assumed that the latent state variable evolves as a first-order Markov Chain with transition probability defined as follows:

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix} \tag{2.27}$$

where p_{lk} represents the probability of moving to state k at time $t+1$ from state l at time t. In our model regime does not affect vine structures $\boldsymbol{\mathcal{V}}_k$ and the set $\mathbf{B}$ of copula families identifying bivariate distributions and conditional bivariate distributions. Therefore, the

MS-V copula is described by its parameters:

$$\boldsymbol{\theta}' = (\boldsymbol{\theta}'_{\boldsymbol{c}}, \boldsymbol{\theta}'_{\boldsymbol{MS}}) = ((\boldsymbol{\theta}'_{\mathbf{1}}, ..., \boldsymbol{\theta}'_{\boldsymbol{p}}), \boldsymbol{\theta}'_{\boldsymbol{MS}})$$

where $\boldsymbol{\theta}_{\boldsymbol{MS}}$ denotes parameters of the transition probability matrix, and $\boldsymbol{\theta}_{\boldsymbol{c}}$ denotes copula parameters. It is also worth noting that the marginal distributions in this specification do not depend on the regime, and hence are modelled separately.

The first challenge in estimating Markov regime-switching models is that the value of S_t is unknown because it is unobservable. Therefore, in order to estimate regime-switching model where the states are not known we can consider a decomposition of the joint density of $\boldsymbol{u}_{1:T} = (\boldsymbol{u}_1, ..., \boldsymbol{u}_T)$:

$$\begin{aligned} c(\boldsymbol{u}_{1:T}|\boldsymbol{\theta}) &= c(\boldsymbol{u}_1|\boldsymbol{\theta}) \cdot \prod_{t=2}^{T} c(\boldsymbol{u}_t|\boldsymbol{u}_{1:(t-1)}, \boldsymbol{\theta}) \\ &= \left[\sum_{k=1}^{p} c(\boldsymbol{u}_t|S_1 = k, \boldsymbol{\theta}_{c,k}) P(S_1 = k|\boldsymbol{\theta}_{\boldsymbol{MS}})\right] \\ &\quad \cdot \prod_{t=2}^{T} \left[\sum_{k=1}^{p} c(\boldsymbol{u}_t|S_t = k, \boldsymbol{\theta}_{c,k}) P(S_t = k|\boldsymbol{u}_{t-1}, \boldsymbol{\theta}_{\boldsymbol{MS}})\right] \end{aligned}$$

Because the required probabilities are not observed, we make inferences on the probabilities at time t using the available information up to period $t-1$. This constitutes the main idea of the filter outlined in Hamilton (1994) and Kim and Nelson (1999), which we use in order to calculate the filtered probabilities of the process being in each state, based on the availability of the new information. This yields the following algorithm that should be iterated through $t = 1, ..., T$:

Step 1: *Prediction of* S_t

$$Pr(S_t = l|w_{t-1}) = \sum_{k=1}^{2} p_{kl} Pr(S_{t-1} = k|w_{t-1})$$

for $l = 1, 2$ and $p_{kl} = Pr(S_t = l|S_{t-1} = k)$ is the transition probability between the states k and l as introduced in (2.27).

Step 2: *Filtering of* S_t

$$Pr(S_t = l|w_t) = \frac{c_t(u_1, u_2|S_t = l, w_{t-1})Pr(S_t = l|w_{t-1})}{\sum_{k=1}^{2} c_t(u_1, u_2|S_t = k, w_{t-1})Pr(S_t = k|w_{t-1})}$$

where $w_t = (w_{t-1}, u_{1t}, u_{2t})'$. At $t = 1$ the filter is initialized using stationary probabilities of S_t:

$\pi_1 = Pr(S_0 = 1|w_0) = \frac{1-p_{22}}{2-p_{11}-p_{22}}$

$\pi_2 = Pr(S_0 = 2|w_0) = \frac{1-p_{11}}{2-p_{11}-p_{22}}$

The smoothed probabilities for S_t $Pr(S_t = l|w_T) = \sum_{k=0}^{1} Pr(S_t = l, S_T = k|w_T)$ can be calculated recursively from the filtered probabilities in the following way as outlined in Hamilton (1994) and Kim and Nelson (1999):

1. Filtered probabilities $Pr(S_t = l|w_t)$ are obtained for $l = 1, 2$ and $t = 2, \ldots, T$.

2. The smoothing algorithm is initialized at time $t = T$ with $Pr(S_T = l|w_T)$.

3. The smoothed probability distribution for each $t = T - 1, \ldots, 1$ is obtained by:

$$Pr(S_t = l|w_T) \quad = \quad Pr(S_t = l|w_t) \sum_{k=1}^{2} \frac{p_{lk} Pr(S_{t+1} = k|w_T)}{\sum_{j=1}^{2} p_{jk} Pr(S_t = j|w_t)} \tag{2.28}$$

where $p_{lk} = Pr(S_{t+1} = k|S_t = l)$ are the transition probabilities between the states l and k.

2.4.2 Extending regime process to a second-order process

The first-order Markov chain model in its original form is known to posses memoryless property. That is, the one-step future evolution of the process is determined by the present state only, and therefore, future and past states are conditionally independent. However, this memoryless property may seem to be inappropriate for time series of financial variables where long memories are present in the data-generating process.

Another restrictive assumption in our first-order Markov-switching model is that transition probabilities are constant. Real world consideration would suggest the plausibility of allowing the transition probabilities to vary with time. It is reasonable to assume that

the longer the economy is in the 'crisis' regime the more likely it is to exit and enter the 'non-crisis' regime. Therefore, it would be sensible not to exclude this possibility from the outset and attempt to capture these dynamics in our model.

We can introduce some degree of flexibility into our model by weakening these restrictive assumptions. One way is to allow regime variable to depend on longer past state sequence. This will allow to incorporate more historical information into our model. This, however, comes with a price of having to estimate extra parameters in the transition probability matrix. Nevertheless, this modelling approach allows to include more states in a highly parsimonious way. In addition, this is a variety of modelling in which the transition parameters depend on how long the process has been in a particular state. For example, in a case of a second-order regime-switching process, the probability that the regime variable will transit into regime 1 given it is currently in regime 2, will depend upon whether the regime variable was already in regime 1 or regime 2 in the last period.

In order to allow the regime random variable to follow a higher-order process, one possibility that suggests itself is to use a model $\{R_t\}$ satisfying the following property for $\ell \geq 2$:

$$P(R_t = i|R_{t-1} = j_{t-1}, R_{t-2} = j_{t-2}, ...) = P(R_t = i|R_{t-1} = j_{t-1}, ..., R_{t-\ell} = k_{t-\ell})$$

Therefore, in a case of a second-order process of regime $\{R_t\}$ we have $\ell = 2$, which satisfies the following property:

$$P(R_t = i|R_{t-1} = j_{t-1}, R_{t-2} = j_{t-2}) \quad \neq \quad P(R_t = i|R_{t-1} = j_{t-1})$$

However, this stochastic process is no longer a Markov chain. We can redefine the model in such a way as to produce a stochastic state process that possesses Markov property. Let $\mathbf{Z}_t = (R_{t-\ell+1}, R_{t-\ell+2}, ..., R_t)$, then $\{\mathbf{Z}_t\}$ is a first-order Markov chain on $\mathcal{S}^\ell$, where $\mathcal{S}$ is the state space of $\{R_t\}$. Therefore, the joint density of $\boldsymbol{u_t}$ and R_t at time t is:

$$c(\boldsymbol{u}_t|R_t = i, R_{t-1} = j, \boldsymbol{\theta}_k)P(R_t = i|\boldsymbol{u}_{1-t}, \boldsymbol{\theta}) \tag{2.29}$$

where $i, j = 1, 2$ and $k = 1, ..., 4$. In order to simplify our model we assume that:

$$c(\boldsymbol{u}_t|R_t = i, R_{t-1} = j, \boldsymbol{\theta}_1) = c(\boldsymbol{u}_t|R_t = i, \boldsymbol{\theta}_1) \tag{2.30}$$

and, therefore, we have that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3 = \boldsymbol{\theta}_4$.

2.4.3 Marginal models

The first step in modelling the dependence structure is to model each marginal distributions separately. This is because most financial returns exhibit some degree of autocorrelation and heteroskedasticity. In order to model marginal distributions of each of our return series we start with initial broad specification searching and allow the data to choose the model for asset return i:

$$\begin{aligned} \phi(L)x_{it} &= \alpha_i + u_{it} \quad \text{where} \quad u_{it} = \sigma_{it}\epsilon_{it} \\ \sigma_{it}^2 &= g(\sigma_{it-1}, \epsilon_{it-1}; \sigma_{it-2}, \epsilon_{it-2}, \ldots) \end{aligned}$$

where $\phi(L)$ is a known lag polynomial denoting finite-order autoregressive model, and $g(\cdot)$ is a known function representing conditional volatility model. Based on AIC and BIC information criteria, the resulting model for the S&P500 returns is the skewed Student's-t GARCH(1,1) expressed as:

$$x_{1t} = \mu_1 + u_{1t} \quad \text{where} \quad u_{1t} = h_{1t}^{1/2} u_{1t} \tag{2.31}$$

$$h_{1t} = \omega_1 + \alpha_1 u_{1t-1}^2 + \beta_1 h_{1t-1} \tag{2.32}$$

where h_{1t} is the conditional variance given past information, which allows capturing persistence, and ϵ_{1t} are *i.i.d.* random variables, ω_1, β_1, $\alpha_1 > 0$ assures that the conditional variance h_{1t} is positive and $\alpha_1 + \beta_1 < 1$ ensures covariance stationarity.

For the FTSE100 and DAX returns the resulting model is the skewed Student-t GJR-

Table 2.4: Descriptive Statistics

	S&P500	**FTSE100**	**DAX**
Mean	0.001369	0.000804	0.001323
Median	0.003178	0.001688	0.003983
Maximum	0.123746	0.166889	0.150603
Minimum	-0.157665	-0.125320	-0.197775
Std. Deviation	0.022907	0.023484	0.031577
Skewness	-0.578704	-0.268820	-0.671595
Kurtosis	7.349300	7.604751	6.384020
Jarque-Bera	1096.356	1163.297	717.4673
	(0.000000)	(0.000000)	(0.000000)

Note: Jarque-Bera corresponds to Jarque-Bera test statistics with p-values in parentheses.

GARCH(1,1) expressed as:

$$x_{it} = \mu_i + u_{it} \quad \text{where} \quad u_{it} = h_{it}^{1/2}\epsilon_{it} \tag{2.33}$$

$$h_{it} = \omega_i + \alpha_i u_{it-1}^2 + \gamma_i u_{it-1}^2 I_{[u_{it-1}<0]} + \beta_i \sigma_{it-1}^2 \tag{2.34}$$

where $i \in \{2 = FTSE100, 3 = DAX\}$. The conditional variance expression now includes the asymmetry term γ, which relaxes the assumption of symmetric response of volatility to positive and negative shocks. This would suggest that in the UK and Germany, a negative shock to equity returns is likely to cause a volatility to rise by more than a positive shock of the same magnitude. In the literature these asymmetries are attributed to *leverage effects*, characterised by an increase in the firm's debt to equity ratio caused by a fall in a firm's stock value.

It is also assumed that the standardised residuals ϵ_{it} for $i = 1, 2, 3$ have a skewed t-distribution. Its density is given by:

$$f(\epsilon_{it}|\upsilon,\lambda) = \begin{cases} bc\left(1 + \frac{1}{\upsilon - 2}\left(\frac{b\epsilon_{it} + a}{1-\lambda}\right)^2\right)^{\frac{-(\upsilon+1)}{2}} & \epsilon_{it} < \frac{-a}{b} \\ bc\left(1 + \frac{1}{\upsilon - 2}\left(\frac{b\epsilon_{it} + a}{1+\lambda}\right)^2\right)^{\frac{-(\upsilon+1)}{2}} & \epsilon_{it} \geq \frac{-a}{b} \end{cases}$$

Table 2.5: Marginal Distributions Estimation Results

	S&P500	FTSE100	DAX
ω_i	0.00001	0.00003	0.00004
	(0.000005)	(0.00001)	(0.00001)
α_i	0.1162	0.0024	0.0356
	(0.0313)	(0.0016)	(0.0174)
γ_i	–	0.2127	0. 1306
	–	(0.0711)	(0.0492)
β_i	0.8665	0.8361	0. 8547
	(0.0352)	(0.0539)	(0.0373)
υ_i	10.8750	9.4166	9.9571
	(3.4733)	(2.51665)	(2.4097)
λ_i	- 0.2696	- 0.1911	-0.2941
	(0.0402)	(0.03699)	(0.0379)
log L	3265.6	3230.4	2848.28
$Q(10)$	0.3016	0.4923	0.4959
$Q^2(20)$	1.0000	1.0000	0.9340
KS	0.9764	0.8838	0.9692
$Berk$	0.9988	0.2620	0.8938

Note: In round brackets are the standard errors. Q and Q^2: Ljung-Box Q statistic for autocorrelation and squared autocorrelation in the residual terms, respectively. Last two rows show goodness-of-fit tests for the probability integral transform of the margins. KS: Kolmogorov - Smirnov test p-value, and $Berk$: Berkowitz test p-value for uniformity.

with constants a, b and c defined as:

$$a = 4\lambda c\left(\frac{\upsilon-2}{\upsilon-1}\right), \quad b^2 = 1+3\lambda^2-a^2, \quad c = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\pi(\upsilon-2)\Gamma\left(\frac{\upsilon}{2}\right)}}$$

where υ correspond to the number of degrees of freedom and λ represents the degree of asymmetry. When λ is negative, we have a left-skewed density, meaning that there is higher probability of observing large negative returns than large positive returns.

Once the residuals from each marginal model are obtained, we standardize them by the corresponding conditional standard deviation. These standardized residuals represent the pseudo-observations with zero-mean, independent and identically distributed series upon which the dependence modelling is based. All three marginal distributions are modelled independent of the regime.

2.5 Data

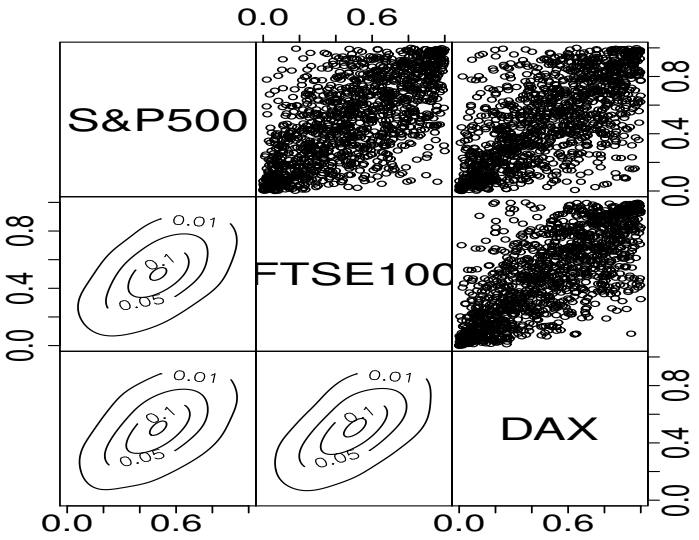
In this study we use weekly data which comprises S&P500, FTSE100 and DAX indices for the period February 1, 1990 to November 18, 2014, which gives us 1299 observations. The data was obtained from DATASTREAM. The stock indices returns are logarithmic weekly returns.

Table 2.4 presents summary statistics of the data. The means for three returns are much smaller relative to their standard deviation. The S&P500 and DAX returns exhibit substantial skewness. The S&P500 and FTSE100 exhibit high kurtosis, whilst DAX exhibits a relatively lower kurtosis. The Jarque-Bera test of normality of three returns series strongly rejects the null, indicative of non-normality of all three series.

2.6 Empirical Results

The estimation results for marginal models are presented in Table 2.5. The asymmetry coefficient is negative and statistically significant for all three series of returns, whilst the coefficient for FTSE100 is relatively small compared to S&P500 and DAX. This indicates that returns series under consideration are negatively skewed, which implies that large negative returns are more likely to occur than large positive returns. The degrees of freedom parameter for all three series is approximately 10, which is indicative of heavy tails in distributions. In addition, the estimate of the β parameter for all three series is approximately 0.8 which suggests high volatility persistence. The significance of the γ parameter for FTSE100 and DAX points towards existence of 'leverage effects', which implies that volatility increases when the stock price falls. The reported high p-values for Kolmogorov-Smirnov (KS) and Berkowitz (Berk) tests for uniformity do not show statistical evidence against the null of Uniform(0,1) distribution of the transformed residuals. Figure 2.9 displays a pair plot of the transformed residuals with scatter plot above the diagonal and empirical contour plots below the diagonal.

Figure 2.9: Transformed standardized residuals



2.6.1 Markov switching vine copula estimation

In this sub-section we present estimation results for different copula family specifications. In order to keep modelling procedure simple, we impose all bivariate blocks to be the same copulas. In addition, it is assumed that the vine structure in low dependence regime is exactly the same as in the high dependence regime. However, the parameters of each bivariate copulas in the vine are allowed to vary across both regimes. Table 2.6 presents estimation results for 10 vine copula candidates with different tail behaviour. The estimation results suggest the presence of two regimes: one with high and one with low dependencies. Based on the likelihood principle, the vine copula with Gaussian copulas as a building block is suggested. Table 2.7 compares this competing model with the first-order regime-switching Gaussian copula. The copula parameter estimates from the high dependence regime do not seem to be statistically different from each other. However, this is not the case for the parameter estimates in the low dependence regime, although the magnitude of the difference does not seem to be immensely large, with the average absolute difference being equal to 7.5.

This table also includes model selection information criteria. Based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), there is more support

Table 2.6: Estimated Copula Parameters

		Regime 1			Regime 2						
Copula	**L**	**12**[a]	**23**	**13|2**	**12**	**23**	**13|2**	$\boldsymbol{P_{11\|1}}$	$\boldsymbol{P_{21\|1}}$	$\boldsymbol{P_{12\|2}}$	$\boldsymbol{P_{22\|2}}$
Gaussian	*1068.7*	0.7719	0.8435	0.4626	0.3735	0.3870	0.1248	0.9859	0.0878	0.3276	0.4312
		(0.0168)	(0.0104)	(0.0344)	(0.0538)	(0.0641)	(0.0574)	(0.1862)	(0.2712)	(0.5625)	(0.7803)
Clayton	*845.29*	10.5975	2.7771	4.6096	1.1174	1.5531	0.3654	0.3799	0.9996	0.0003	0.0165
Survival Gumbel	*954.02*	5.4250	3.8313	2.7959	1.6989	1.9496	1.2462	0.7952	0.0267	0.0230	0.0796
Frank	*1026.78*	6.9637	8.5662	3.1961	2.6673	3.2620	0.9411	0.9927	0.0600	0.0498	0.3713
Joe	*707.69*	2.4602	2.7937	1.3881	1.1234	1.0746	1.0189	0.9901	0.0492	0.3205	0.8546
Student t	*1055.87*	0.7629	0.8323	0.4420	0.4622	0.4939	0.1861	0.9999	0.0073	0.6054	0.1902
	–	20.7269	20.3785	20.3270	22.0110	11.9366	13.8142	–	–	–	–
BB1	*1045.08*	0.33607	0.6515	0.4175	0.6634	0.8951	0.0903	0.9999	0.0087	0.7291	0.1003
	–	1.8150	1.8979	1.1525	1.0896	1.0342	1.0709	–	–	–	–
BB6	*825.02*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9927	0.0361	0.3954	0.9999
	–	2.0874	2.4143	1.3238	1.1490	1.1217	1.1000	–	–	–	–
BB7	*910.03*	1.9478	2.2668	1.1024	1.0663	1.0000	1.0000	0.9892	0.4644	0.0508	0.0387
	–	1.0000	1.3558	1.0000	1.0000	1.0000	1.0000	–	–	–	–
BB8	*949.78*	7.1949	8.5332	5.1215	3.2627	3.9729	2.0418	0.9937	0.0468	0.1434	0.6324
	–	0.6276	0.6306	0.5051	0.5493	0.5039	0.4367	–	–	–	–

[a]Variable 1 = S&P500, 2 = FTSE100, 3 = DAX. Standard errors are reported only for the "best" model due to computational intensity. Computation of standard errors is described in Appendix A1.

Table 2.7: Second-order regime switching copula estimation results

	High Regime		**Low Regime**	
	2nd order	1st order	2nd order	1st order
S&P500/FTSE100	0.7719	0.7669	0.3735	0.4556
	(0.0168)		(0.0538)	
FTSE100/ DAX	0.8435	0.8411	0.3870	0.4716
	(0.0104)		(0.0641)	
S&P500/DAX	0.4626	0.4536	0.1248	0.1854
	(0.0344)		(0.0574)	

Transition probabilities

p1	**p2**	**p3**	**p4**
0.9856	0.0878	0.3284	0.4312
(0.1862)	(0.2712)	(0.5625)	(0.7803)

Model Selection

Model	**LL**	**AIC**	**BIC**
2nd order	1068.7	-2117.4	-2065.7
1st order	1059.6	-2103.4	-2061.8

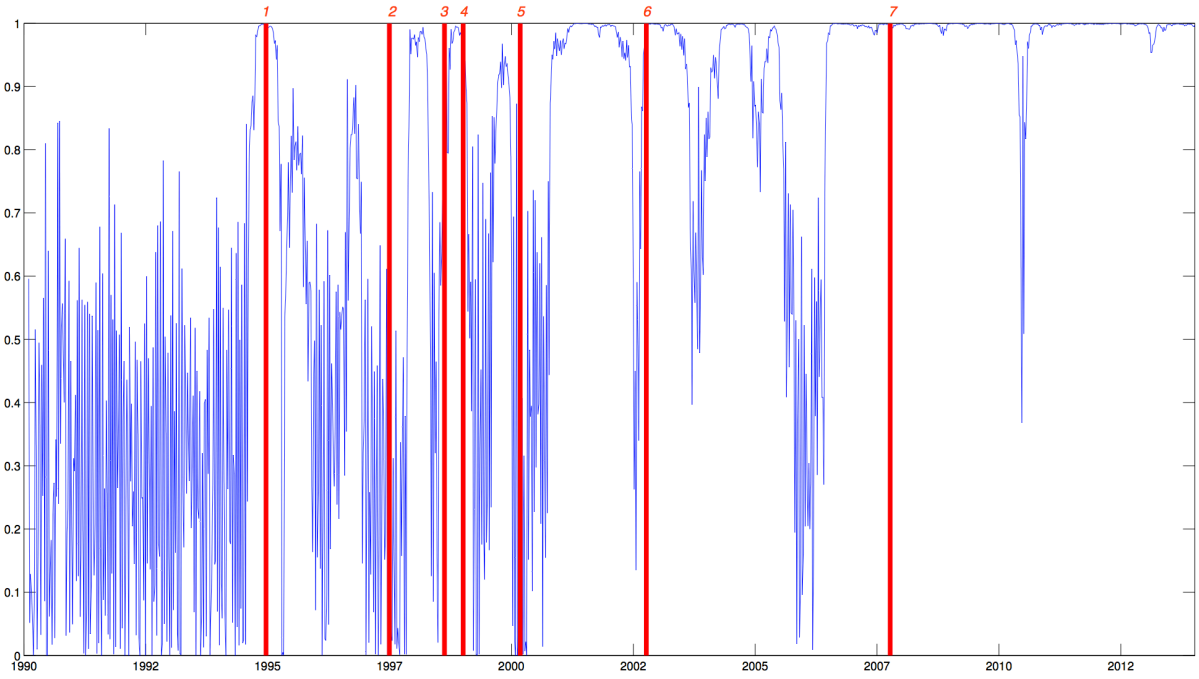
Note: In round brackets are the standard errors for the parameters of the second-order regime switching model. Parameter estimates for the first-order regime switching model have been included for comparison only, and hence their standard errors have not been computed.

for the second-order regime-switching model, even with extra 2 parameters[3]. Hence, the preferred model is the second-order regime-switching vine copula.

Figure 2.10 shows the plot of the smoothed probability of being in high dependence regime. This is plotted against time, and a vertical axis having a unit length. The vertical red lines represent economic and financial crises that took place throughout the world during our sample period: starting with economic crisis in Mexico in 1990; Asian financial crisis of 1997; followed by the Russian financial crisis in 1998; ensuing Argentine economics crisis in 1998; the Dot-com bubble in early 2000s; the sharp drop in stock prices during 2002 in stock markets across the US, Canada, Asia, and Europe; and the financial crisis of 2007. From the graph it can be observed that most of the time the increase in probability of being in high dependence regime coincides with the financial crises throughout the world. This lends additional support to the empirical findings in the literature that the strength of dependence increases during financial crises.

[3]The model-selection procedure is based on choosing the model for which the AIC and BIC is the smallest.

Figure 2.10: Smoothed Probability of being in HIGH regime (2nd order regime process)



1 = Economic crisis in Mexico; 2 = Asian financial crisis; 3 = Russian financial crisis; 4 = Argentine economics crisis; 5 = Dot-com bubble; 6 = sharp drop in stock prices across the US, Canada, Asia, and Europe; 7 = Financial crisis of 2007.

2.6.2 Dependence Measures

In order to relate results and compare the implied dependence strength by various copulas in each regime, we can reparameterize copula's parameter in terms of copula-based Blomqvist's β measure of dependence. The Blomqvist's β is defined as follows:

$$\beta = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1 \tag{2.35}$$

There are other copula-based measures of dependence such as Kendall's τ and Spearman's ρ that are invariant under strictly increasing transformations. However, only Blomqvist's β will be used as it is computationally simpler to compute relative to above mentioned measures of dependence. Table 2.8 presents estimates of Blomqvist's betas for all bivariate copulas in both regimes. Blomqvist's betas computed for conditional copulas in the lower trees are not straightforward to interpret, and, therefore, we only concentrate our attention on the unconditional bivariate copulas in the first tree. Looking at the dependence between S&P500 and FTSE100 the strongest dependence is implied by regime-switching Clayton

Table 2.8: Blomqvist β estimates

		Gaussian	Clayton	Survival Gumbel	Frank	Joe	Student's t	BB1	BB6	BB7	BB8
12	Regime 1	0.56	0.87	0.82	0.62	0.45	0.55	0.52	0.52	0.46	0.58
	Regime 2	0.24	0.36	0.41	0.31	0.06	0.31	0.30	0.13	0.34	0.24
23	Regime 1	0.64	0.60	0.74	0.68	0.50	0.63	0.59	0.59	0.53	0.64
	Regime 2	0.25	0.44	0.49	0.37	0.04	0.33	0.33	0.11	0.33	0.27
13|2	Regime 1	0.31	0.73	0.65	0.36	0.17	0.29	0.27	0.24	0.35	0.35
	Regime 2	0.08	0.15	0.19	0.12	0.01	0.12	0.10	0.09	0.33	0.08

vine copula with Blomqvist's beta being 0.87 in high dependence regime and 0.36 in low dependence regime [4]. This constitutes an increase in the strength of dependence by more than 100 per cent when we move into 'crisis' regime. Similarly, the weakest dependence is implied by Joe copula which amounts to 0.45 in the high dependence regime and 0.06 in the low dependence regime. The strongest dependence between FTSE100 and DAX is implied by the Survival Gumbel copula which is 0.74 in the high dependence regime and 0.49 in the low dependence regime. For this pair of indices the weakest dependence is implied by Joe copula with a Blomqvist's beta being 0.5 in the high dependence regime and 0.04 in the low dependence regime.

2.7 Conclusion

In this paper Markov switching vine copula is extended to a second-order regime switching vine copula. This modelling approach identified two regimes: one of high and one of low dependence between international stock indices returns. In order to allow asymmetries in the upper and lower tail-dependence, we have compared 10 vine copulas with different tail behaviour. For the sample period under investigation, based on the principle of maximum likelihood, Gaussian bivariate copulas were selected as main building blocks in the vine copula specification. We also compared the selected model against the first-order regime switching vine copula. The information criteria used to aid model selection procedure suggests that the second-order regime switching vine copula may be an adequate choice for the sample under study.

This finding seems to indicate the absence of heavy tails and tail asymmetry. In addi-

[4]Clayton vine copula is a vine copula in which all bivariate blocks consist of Clayton copulas. This definition also applies to other vine copulas.

tion, this specification implies that the downside risk vanishes asymptotically. In portfolio risk management, this would indicate absence of non-linear portfolio dependence. However, our conclusion should not be solely based on the likelihood principle. It cannot at this stage be conclusively judged which copula fits best the data at hand. It well may be that some copulas fit better in the middle, whilst other fit better near the lower or upper tail, or even both. Therefore, the fit in the tails should also be taken into consideration.

Chapter 3

Analysing asymptotic sequences of Markov regime-switching Gaussian copula processes local to absorbing states

3.1 Introduction

The standard procedure in the literature is to rely on the asymptotic theory based on the assumption that the probability transition matrix remains fixed as the sample size increases to infinity. This essentially means that as the sample gets increasingly large, the fraction of time spent in both states settles down to their long-run proportions. Therefore, the time spent in each state is non-negligible, and the parameter estimates in both regimes are well behaved. In the alternative asymptotic sequence we might be concerned that we could spend most of the time in one regime rather than in the other regime. Hence, the asymptotic sequence in which the fraction of the time spent in one of the states converges to something moderate may not be informative about the finite sample distributions. The asymptotic theory that is generally used would not be appropriate under these circumstances. This state of the world, where recessions last only a short period of time, would be more relevant and interesting to investigate. The purpose of this chapter is to conduct an investigation through a Monte Carlo study in order to examine under what conditions models of this variety work well, and if there are systematic ways in which they tend to not work so well.

Analysing sequences of processes is a well-established technique. For example, a sequence of local alternatives is examined in the analysis of the asymptotic power of a consistent test. The sequences of processes have also been considered in the study of nearly non-stationary AR(1) processes. For example, Chan and Wei (1987) studied the limiting distribution of the least-squares estimate of the coefficient in the AR(1) model. In another study, Andrews (2001) considered testing problems when parameters in the null hypothesis lie on the boundary of the maintained hypothesis space. The author provides general asymptotic results that take these testing problems into account.

This chapter is structured as follows. Section 3.2 provides a brief introduction to a Markov Chain theory, and also a mathematical description of the two alternative asymptotic sequences that will be analysed in this study. Section 3.3 reports and discusses results from a Monte Carlo study. Section 3.5 provides concluding discussion.

3.2 Framework for alternative sequences

The situation we are considering is local to the limit of the transition probability matrix with one or more absorbing states. First, we present briefly a theory of Markov chains where the transition probability matrix is fixed. We consider a discrete-time Markov chain in which the state can change at each discrete time point. Given a sequence of discrete random variables $S_0, S_1, S_2, ...$ taking values in some finite or countably infinite set Ω, a sequence S_n is said to be a Markov chain if the following Markov property is satisfied:

$$P(S_{n+1} = j | S_n = i, S_{n-1} = i_{n-1}, ..., S_0 = i_0) = P(S_{n+1} = j | S_n = i) \tag{3.36}$$

We can therefore define one-step transition probabilities for a two-state Markov process as follows:

$$p_{ij} = P(S_{n+1} = j | S_n = i) \quad \forall i, j = 1, 2 \tag{3.37}$$

where S_t denotes the state at date t. We can then collect these probabilities to construct

the transition probability matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

where $p_{ij} \geq 0$ for all $i, j = 1, 2$ so that $p_{11} + p_{12} = 1$ and $p_{21} + p_{22} = 1$

The basic model of interest is a regime-switching copula model. We use bivariate time series generated from a regime-switching Gaussian copula model, where the true data generating process is specified to be:

$$c(\boldsymbol{u_t}|\boldsymbol{\Theta}, S_t) = \sum_{k=1}^{2} \mathbb{1}_{\{k\}}(S_t) \cdot c(\boldsymbol{u_t}|\boldsymbol{\theta}_k) \tag{3.37}$$

where $\boldsymbol{\Theta}$ is a vector of population parameters which includes parameter vectors $\boldsymbol{\theta}_1$ in *state 1*, and $\boldsymbol{\theta}_2$ in *state 2*; S_t is the latent state variable defined in (3.37); $c(\boldsymbol{u_t}|\boldsymbol{\theta}_k)$ is a bivariate Gaussian copula which has the following form:

$$c(u_{1t}, u_{2t}|\boldsymbol{\theta}_k) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{(\Phi^{-1}(u_1)^2 - 2\rho\Phi^{-1}(u_1)\Phi^{-1}(u_2) + \Phi^{-1}(u_2)^2)}{2(1-\rho^2)}\right\}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal, and $\rho \in (-1, 1)$.

This specification implies that the copula density of a vector $\boldsymbol{u_t}$ at time t depends on a random variable S_t that indicates the current regime. In this specification, the Gaussian copula is chosen for both regimes.

The general framework that is being considered is a sequence of processes $\{\mathcal{M}^{(n)}\}_{n=1}^{\infty}$ associated with a sequence of run lengths $\{T^{(n)}\}_{n=1}^{\infty}$, with accompanying transition matrices $\left\{p_{ij}^{(n)}\right\}_{n=1}^{\infty}$ and initial draw distributions $\left\{\pi_j^{(n)}\right\}_{n=1}^{\infty}$, where each $\pi_j^{(n)}$ is determined by $p_{ij}^{(n)}$. In this sequence, the copula parameters in each state remain the same. Given a finite value of n, the probability transition matrix is well-behaved[5]. Nevertheless, as n increases to infinity, the probability transition matrix converges to a limit matrix, which will be

[5]Here "well-behaved" means that all states can be reached from any other state with a non-negligible probability.

defined in the next section.

3.2.1 Alternative asymptotic sequences

In this subsection we consider two alternative asymptotic sequences. In particular, we consider a limit of sequences of transition matrices. This is about situations where we are local to a limit as a consequence of considering a sequence of processes where each process is indexed by the run length. Consequently, the higher the run length is, the closer we are to the limit. In this framework, the general asymptotic theory is discontinuous in the transition probability parameter.

Case 1

The first asymptotic sequence involves a transition matrix converging to a matrix in which all elements of the first column are equal to one, and all elements of the second column are equal to zero:

$$\lim_{T\to\infty} P_T = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \tag{3.38}$$

The implications of this is that, as the run length T increases, the chain will tend to spend most of the time in state 1 and, some of the time in state 2.

We can define transition matrix P as a function of G, M and T:

$$P_T = I - M - T^{-1}G$$

where

$$M = \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} \qquad \text{and} \qquad G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} 100 & -100 \\ 100 & -100 \end{pmatrix}$$

Using this construction, we can find the corresponding probability transition matrix P_T for any T:

$$P_T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} - T^{-1}\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

$$= \begin{pmatrix} (1 - T^{-1}g_{11}) & -T^{-1}g_{12} \\ (1 - T^{-1}g_{21}) & -T^{-1}g_{22} \end{pmatrix} \tag{3.39}$$

In this specification it is evident that, in order for the transition probabilities to be positive and lie between 0 and 1, the following restrictions have to be placed on elements of the G matrix [6]:

1. $g_{11} > 0$, $g_{21} > 0$, $g_{12} < 0$ and $g_{22} < 0$
2. $g_{11} + g_{12} = g_{21} + g_{22} = 0$

The ergodic distribution can also be derived as follows:

$$\pi_1^T = \frac{1 - (-T^{-1}g_{22})}{[1 - (1 - T^{-1}g_{11})] + [1 - (-T^{-1}g_{22})]} = \frac{T + g_{22}}{T + g_{11} + g_{22}} \tag{3.40}$$

$$\pi_2^T = \frac{1 - (1 - T^{-1}g_{11})}{[1 - (1 - T^{-1}g_{11})] + [1 - (-T^{-1}g_{22})]} = \frac{g_{11}}{T + g_{11} + g_{22}} \tag{3.41}$$

Evidently, ergodic probabilities depend on T with the following limits:

$$\begin{aligned} \lim_{T\to\infty} \pi_1^T &= 1 \\ \lim_{T\to\infty} \pi_2^T &= 0 \end{aligned}$$

We can also derive the expected number of state transitions by chain between date 0 and date T:

$$N_T^e = \frac{2T[1 - (1 - T^{-1}g_{11})][1 - (-T^{-1}g_{22})]}{[1 - (1 - T^{-1}g_{11})] + [1 - (-T^{-1}g_{22})]} = \frac{T2g_{11} + 2g_{11}g_{22}}{T + g_{11} + g_{22}} \tag{3.42}$$

Similarly, it is evident that the expected number of state transitions also depends on T with the following limit:

$$\lim_{T\to\infty} N_T^e = 2g_{11} \tag{3.43}$$

We can also derive the expected number of transitions from *state 2* to *state 2* for a given

[6] These restrictions are still not sufficient to ensure that the elements of the transition probability matrix P_T are non-negative. There exist certain combinations of run length T and matrix G (which we do not consider in this study) such that the resulting P_T matrix is negative. This is, however, not a serious issue, as there is a simple modification to matrix P_T that ensures non-negativity. More details on the procedure can be found in Appendix A.2.1.

run length T, which we denote as $\mathbb{E}\left(n_{22}^{T}\right)$. There is only one way in which the Markov chain can have a transition from *state 2* to *state 2* at date t. In particular, at date *(t -1)* the chain is in *state 2*, and then in the next time period t it remains in *state 2*. Using (3.41) and (3.39), we can derive the unconditional probability of this event happening for a given run length T:

$$\begin{aligned} Pr\left(S_{t-1}^{T}=2,\ S_{t}^{T}=2\right) &= Pr\left(S_{t}^{T}=2 \mid S_{t-1}^{T}=2\right) Pr\left(S_{t-1}^{T}=2\right) \\ &= p_{22}^{T}\,\pi_{2}^{T} \\ &= -\frac{g_{22}}{T}\frac{g_{11}}{T+g_{11}+g_{22}} \\ &= -\frac{g_{22}\,g_{11}\,T^{-1}}{T+g_{11}+g_{22}} \end{aligned} \tag{3.44}$$

The expected number of state transitions from *state 2* to *state 2* between date *0* and date T is simply:

$$\mathbb{E}\left(n_{22}^{T}\right) = T\,p_{22}^{T}\,\pi_{2}^{T} = \frac{-g_{22}\,g_{11}}{T+g_{11}+g_{22}} \tag{3.45}$$

which tends to *0* as T tends to infinity, $\lim_{T\to\infty}\mathbb{E}\left(n_{22}^{T}\right)=0$.

Case 2

The second case concerns a scenario where the transition matrix approaches an identity matrix I as $T\to\infty$:

$$\lim_{T\to\infty} P_{T} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.46}$$

The implication of this is that, as the run length T increases, the chain will either tend to spend most of the time in *state 1* or, most of the time it will spend in *state 2*.

We begin by defining P as a function of G and T:

$$P_{T} = I - T^{-1}G$$

where G is a square matrix:

$$G = \begin{pmatrix} g_{11} & -g_{12} \\ -g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} 100 & -100 \\ -900 & 900 \end{pmatrix}$$

To insure ergodicity, the following restrictions have to be placed on elements of the G matrix[7]:

1. $g_{11} > 0$, $g_{12} < 0$, $g_{21} > 0$ and $g_{22} > 0$
2. $g_{11} + g_{12} = g_{21} + g_{22} = 0$

Therefore, we can easily find the structure of P_T for any T:

$$P_T = \begin{pmatrix} (1 - T^{-1}g_{11}) & -T^{-1}g_{12} \\ -T^{-1}g_{21} & (1 - T^{-1}g_{22}) \end{pmatrix} \tag{3.47}$$

The usefulness of this expression is that we can derive the ergodic distribution for chain given T as follows:

$$\pi_{1,T} = \frac{1 - (1 - T^{-1}g_{22})}{[1 - (1 - T^{-1}g_{22})] + [1 - (1 - T^{-1}g_{11})]} = \frac{g_{22}}{g_{11} + g_{22}} \tag{3.48}$$

$$\pi_{2,T} = \frac{1 - (1 - T^{-1}g_{11})}{[1 - (1 - T^{-1}g_{22})] + [1 - (1 - T^{-1}g_{11})]} = \frac{g_{11}}{g_{11} + g_{22}} \tag{3.49}$$

We can also derive the expected number of state transitions by chain given T:

$$N_T^e = \frac{2T[1 - (1 - T^{-1}g_{22})][1 - (1 - T^{-1}g_{11})]}{[1 - (1 - T^{-1}g_{22})] + [1 - (1 - T^{-1}g_{11})]} = \frac{2g_{22}g_{11}}{g_{11} + g_{22}} \tag{3.50}$$

This specification allows to explore what happens when the transition matrix is close to the identity matrix.

Similarly, we can derive the expected number of transitions from *state 1* to *state 1* for a given run length T, denoted as $\mathbb{E}\left(n_{11}^T\right)$. As it was discussed previously, there is only one way in which the Markov chain can have a transition from *state 1* to *state 1* at date *t*. In particular, at date *(t -1)* the chain is in *state 1*, and then in the next time period t it remains in *state 1*. Using (3.47) and (3.49), we can derive the unconditional probability of

[7] See comments in footnote 6 on page 49.

this event happening for a given run length T:

$$\begin{aligned} Pr\left(S_{t-1}^T = 1,\ S_t^T = 1\right) &= Pr\left(S_t^T = 1 \mid S_{t-1}^T = 1\right) Pr\left(S_{t-1}^T = 1\right) \\ &= p_{11}^T \pi_1^T \\ &= \frac{(1 - g_{11} T^{-1}) g_{22}}{g_{11} + g_{22}} \end{aligned}$$

The expected number of state transitions from *state 1* to *state 1* between date *0* and date T is simply:

$$\begin{aligned} \mathbb{E}\left(n_{11}^T\right) &= T p_{11}^T \pi_1^T \\ &= \frac{(T - g_{11}) g_{22}}{g_{11} + g_{22}} \end{aligned} \tag{3.51}$$

which tends to infinity as T tends to infinity, $\lim_{T \to \infty} \mathbb{E}\left(n_{11}^T\right) = \infty$.

In a similar vein, we can also derive the expected number of state transitions from *state 2* to *state 2* for a given run length T:

$$\begin{aligned} \mathbb{E}\left(n_{22}^T\right) &= T p_{22}^T \pi_2^T \\ &= \frac{(T - g_{22}) g_{11}}{g_{11} + g_{22}} \end{aligned} \tag{3.52}$$

which also increases to infinity as T tends to infinity, $\lim_{T \to \infty} \mathbb{E}\left(n_{22}^T\right) = \infty$.

3.3 Monte Carlo simulations

In this section, we study the behaviour of the standardised sampling error in each element of the vector of parameter estimates $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{p}_{11}, \widehat{p}_{21})'$ from the Markov regime-switching copula model in (3.37). The estimated density plots of their sampling distributions from a set of Monte Carlo simulations are provided to illustrate their finite sample distributions.

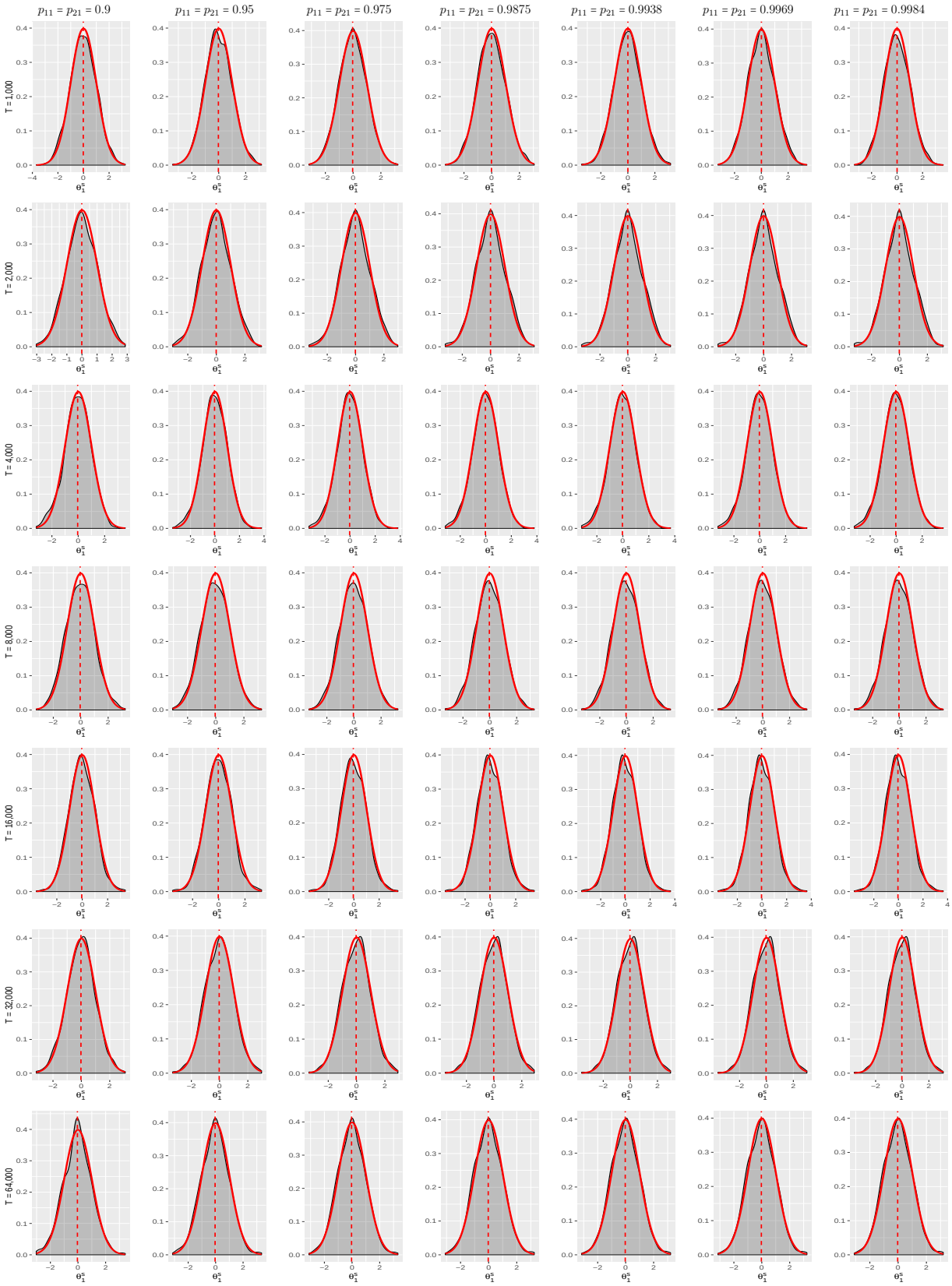
3.3.1 Alternative asymptotic sequence Case 1

First we report results for the Case 1 of the alternative asymptotic sequence described by specification in (3.38) and (3.39). In this scenario the transition matrix parameter

Table 3.9: Summary statistics for the rescaled sampling error $\theta_1^s = (\hat{\theta}_1 - \theta_1)/se(\hat{\theta}_1)$ when states are observed.

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	-0.0259	-0.0228	-0.022	-0.0235	-0.0228	-0.0208	-0.0211
Median	-0.0441	-0.0471	-0.0427	-0.0463	-0.0606	-0.0435	-0.0474
Interquartile	1.3783	1.3718	1.3742	1.3901	1.4072	1.4094	1.4183
Std	1.0217	1.0081	1.008	1.0077	1.0077	1.0086	1.0103
Skewness	0.0208	0.0143	0.0319	0.0365	0.038	0.0351	0.0309
Kurtosis	2.8379	2.8009	2.7629	2.71	2.7171	2.7032	2.7063
Min	-3.0392	-3.1854	-3.1802	-3.2139	-3.2775	-3.2366	-3.351
Max	2.9356	2.6451	2.7173	2.6018	2.5141	2.5284	2.5421
T=2,000							
Mean	-0.0128	-0.0108	-0.0139	-0.0101	-0.0122	-0.0102	-0.0098
Median	-0.0249	-0.0094	-0.008	-0.0049	-0.0256	-0.0146	-0.0176
Interquartile	1.4135	1.3941	1.3813	1.3909	1.3877	1.3622	1.3525
Std	1.0224	1.0194	1.0259	1.0254	1.0281	1.0288	1.0281
Skewness	-0.0031	-0.0035	-0.0436	-0.0565	-0.0571	-0.0516	-0.0535
Kurtosis	2.8333	2.8531	2.9423	2.944	2.9795	2.9877	3.0026
Min	-3.0294	-3.0604	-3.2914	-3.1995	-3.3214	-3.2734	-3.2341
Max	2.8661	3.1613	2.9844	3.0703	3.0618	3.0949	3.1337
T=4,000							
Mean	-0.0436	-0.0353	-0.0349	-0.0294	-0.0317	-0.0298	-0.0295
Median	-0.0285	-0.0281	-0.0153	-0.0227	-0.0436	-0.0382	-0.0401
Interquartile	1.3196	1.3431	1.3208	1.31	1.3192	1.3313	1.3273
Std	1.0076	1.0075	1.0105	1.0062	1.0066	1.0073	1.0069
Skewness	-0.1627	-0.1401	-0.125	-0.0961	-0.0975	-0.0921	-0.0988
Kurtosis	3.0525	3.1515	3.1742	3.1162	3.1464	3.1517	3.15
Min	-3.1814	-3.4569	-3.2754	-3.1684	-3.1746	-3.2373	-3.2186
Max	3.5566	3.796	3.8065	3.7875	3.6728	3.6469	3.6125
T=8,000							
Mean	-0.0414	-0.0379	-0.0399	-0.0329	-0.0314	-0.032	-0.0312
Median	-0.058	-0.0409	-0.0412	-0.0497	-0.0585	-0.056	-0.0439
Interquartile	1.3972	1.3998	1.3983	1.4022	1.3985	1.3794	1.3621
Std	1.0241	1.0181	1.0269	1.0269	1.0277	1.0294	1.0288
Skewness	0.0825	0.0251	-0.0052	0.0008	0.0035	0.0029	0.0015
Kurtosis	2.9428	2.8849	2.8838	2.927	2.9282	2.9458	2.9442
Min	-3.2448	-3.0608	-3.3185	-3.3139	-3.3585	-3.3624	-3.3809
Max	3.2442	3.2508	3.2375	3.3039	3.2998	3.2941	3.289
T=16,000							
Mean	-0.0392	-0.038	-0.046	-0.036	-0.0374	-0.0367	-0.0372
Median	-0.0457	-0.047	-0.0848	-0.0721	-0.0761	-0.0693	-0.0735
Interquartile	1.3374	1.3702	1.3681	1.3636	1.3166	1.3528	1.3444
Std	1.0028	0.998	0.9998	0.9975	0.9992	1.0008	1.0005
Skewness	0.0715	0.0367	0.0664	0.0667	0.0774	0.0817	0.0804
Kurtosis	3.208	3.235	3.2251	3.2508	3.2941	3.2962	3.2832
Min	-3.6168	-3.5249	-3.5667	-3.5652	-3.5471	-3.5748	-3.588
Max	3.3799	3.2643	3.4862	3.4623	3.6523	3.6423	3.6427
T=32,000							
Mean	-0.0285	-0.0273	-0.0302	-0.0242	-0.0282	-0.0288	-0.0294
Median	0.0074	-0.0104	-0.0041	-0.0133	-0.0056	-0.0012	-0.0172
Interquartile	1.3027	1.3592	1.3345	1.3326	1.3369	1.3375	1.3289
Std	1.0004	0.9972	0.9925	0.9919	0.9919	0.9928	0.9923
Skewness	-0.0359	-0.0152	-0.0217	-0.02	-0.0207	-0.0195	-0.0209
Kurtosis	3.0672	3.0471	3.0494	3.0709	3.1088	3.1117	3.1263
Min	-3.1782	-3.3269	-3.3651	-3.524	-3.5962	-3.6084	-3.6021
Max	3.0913	2.9709	2.9165	2.8973	2.9642	3.0047	3.0745
T=64,000							
Mean	-0.0398	-0.0398	-0.0355	-0.0298	-0.0294	-0.0303	-0.0301
Median	-0.0377	-0.0372	-0.0211	-0.0149	-0.0222	-0.0119	-0.0038
Interquartile	1.3578	1.3532	1.385	1.395	1.4044	1.3889	1.3939
Std	1.0173	1.008	1.0008	1.002	0.9999	1.0021	1.0018
Skewness	0.007	0.0136	0.0455	0.0582	0.0384	0.0468	0.0434
Kurtosis	3.1573	3.109	3.0346	3.0331	3.0091	3.0033	3.0046
Min	-3.0212	-3.1341	-3.1186	-3.0878	-3.1479	-3.1423	-3.098
Max	3.4186	3.3441	3.3134	3.2412	3.2078	3.2313	3.1876

Figure 3.11: Distribution of the rescaled sampling error $\theta_1^s = (\hat{\theta}_1 - \theta_1)/se(\hat{\theta}_1)$ when states are observed.



Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000

values change with the run length in a fashion that results in the expected number of time periods spent in one state being constant rather than proportional to the run length.

Observed states

The standard analysis would suggest that if we could observe the state we were in, this would be playing an important role in the estimation of parameters of interest. If one could observe the states, then effectively the entire data could be split into *state 1* data and *state 2* data. One would have one amount of observations in *state 1* and one amount of observations in *state 2*, and this would greatly simplify the estimation procedure of parameters of interest. In what follows, we present the simulations results when states are fully observed, and comment on the limiting behaviour.

The rescaled sampling error in copula parameter θ_1 in state 1

Let's consider a world in which one could observe the states. Figure 3.11 shows the sampling distribution of $\theta_1^s = (\hat{\theta}_1 - 0.8)/se(\hat{\theta}_1)$ in the high occupancy state. Along the main diagonal we can observe the behaviour of the rescaled sampling error under the alternative asymptotic sequence Case 1 described in (3.38). Under this sequence, the run length T and the transition probability matrix P are allowed to vary, whilst the G matrix is held constant.

The behaviour of the rescaled sampling error under the usual asymptotic sequence can be examined by moving vertically down the column. Under this sequence, the P matrix is held constant whilst the matrix G and the run length T are allowed vary. Moving across columns along each row of the Figure 3.11, one can also analyse how behaviour changes whilst holding the run length fixed at a particular T, and varying matrices P and G. These combinations can provide a good sense of how would things progress if matrix G would get close to zero, or away from zero, and at what point noticeable issues would start to arise.

When states can be observed, one knows with absolute certainty whether the stochastic system was in *state 1* or *state 2*, and hence the copula parameter estimates do not depend as such on transition parameters p_{11} and p_{21}. They solely depend on the values of copula pair when the stochastic system was in *state 1* or *state 2*. In other words, the data could

be effectively split into whether it was generated from *state 1* or *state 2*. As one goes down the main diagonal, the heuristic theory would suggest that the rescaled distribution of the sampling error would converge to the standard normal distribution, which seems to be the case when conducting a visual inspection of the plots. For example, as one moves down the main diagonal by one cell, the run length T doubles, and the number of time periods spent in *state 1* more than doubles by the fixed amount equal to g_{11}. In essence, one would get more information about the *state 1* than under the usual asymptotic sequence for a given increase in the sample size, whilst remaining away from the boundary of the parameter space. Consequently, the expected amount of time spent in *state 1* will increase to infinity as the run length T increases to infinity. Hence, moving sufficiently far down the asymptotic sequence, one has a large enough sample from *state 1*, and therefore copula parameters from the high occupancy *state 1* are being well determined most of the time, and the usual asymptotics seem to work well most of the time. The behaviour under the usual asymptotics can be analysed by moving vertically down the columns. In this standard asymptotic sequence, T and G vary whilst matrix P is being held fixed. This results in the expected number of time periods spent in *state 1* to double as the run length doubles.

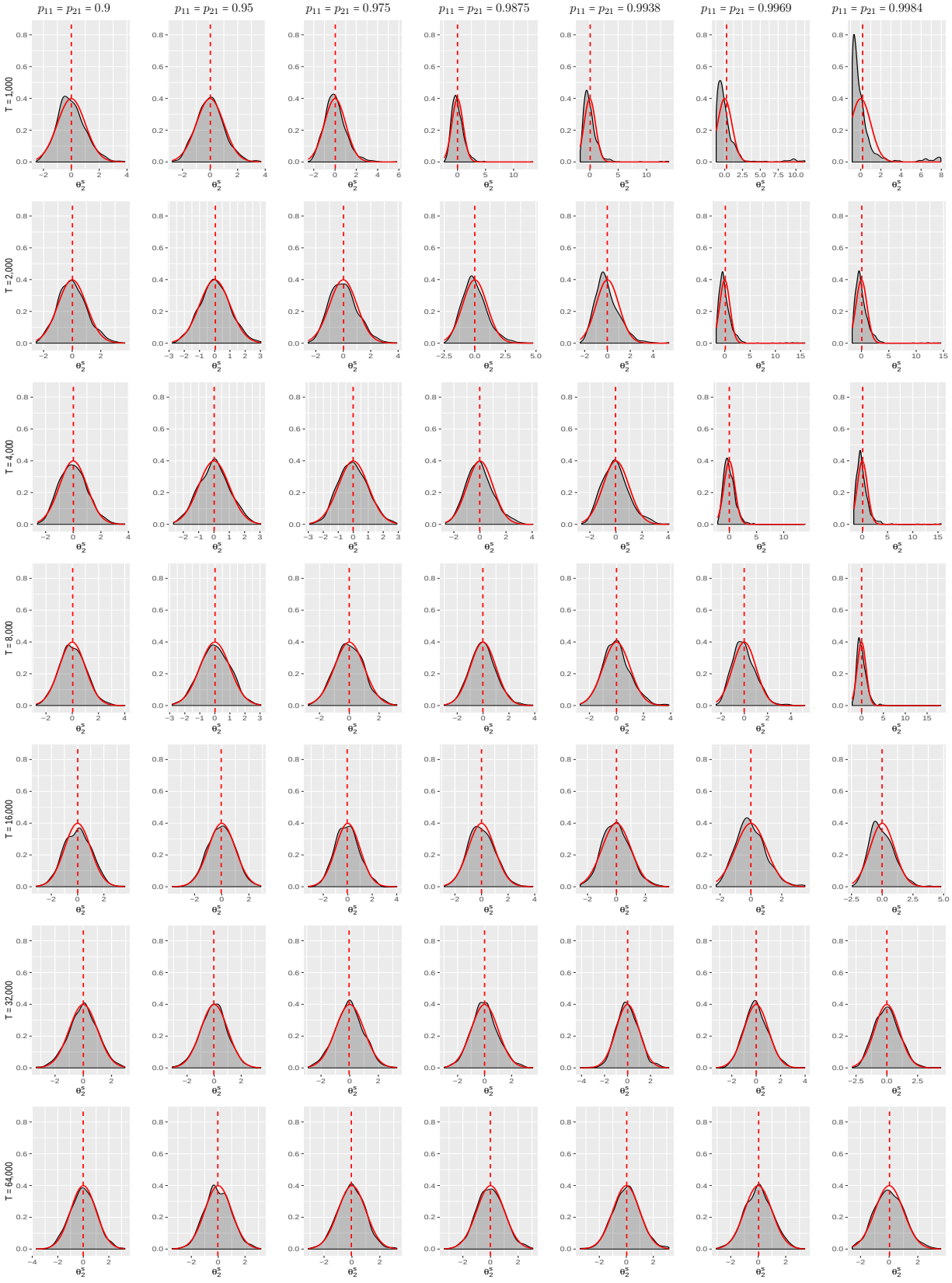
We could also analyse what happens as one moves along one of the rows across columns. This would imply that for a given fixed T the transition probability matrix P and matrix G would vary, resulting in the number of time periods spent in *state 1* to increase by g_{11}. As a result, one would obtain additional information about the copula parameter in *state 1*, and hence one would expect sampling behaviour to improve. Evidently, the amount of additional information one would obtain would depend on how far away from zero the G matrix is. The superimposed reference standard normal distribution makes any discrepancies easily comparable, and the plots seem to indicate that there are no noticeable discrepancies in all cells. Furthermore, from the summary statistics presented in Table 3.9 it can be seen that the interquartile range, standard error, skewness and kurtosis, all remain close to the reference distribution. The rate at which the sampling distribution is shrinking seems to be broadly in line with what the heuristic theory would suggest, that is at the rate

Table 3.10: Summary statistics for the rescaled sampling error $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$ when states are observed.[1]

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	0.0022	0.0192	0.0228	0.0173	0.0671	0.2529	0.1849
Median	-0.0805	-0.0044	-0.0706	-0.1361	-0.2278	-0.2682	-0.3198
Interquartile	1.289	1.2895	1.2399	1.3441	1.2798	1.2231	0.9326
Std	0.9928	0.9861	1.0114	1.1012	1.4295	1.9918	1.6242
Skewness	0.4497	0.3106	0.7512	2.4241	4.5379	3.6487	3.3254
Kurtosis	3.2786	3.2613	4.5423	25.0647	39.0807	17.6538	14.5257
Min	-2.5225	-2.7457	-2.5205	-2.3596	-1.6897	-1.1948	-0.8449
Max	3.847	3.6843	5.8109	13.4444	13.9418	11.2563	7.9594
T=2,000							
Mean	0.0165	0.0375	0.0204	-0.003	-0.0158	0.1133	0.0650
Median	-0.0288	-0.0041	-0.0453	-0.1299	-0.1718	-0.2271	-0.2288
Interquartile	1.3623	1.3593	1.3822	1.3308	1.3092	1.3569	1.3271
Std	1.0085	0.9961	1.0132	1.028	1.0575	1.6964	1.4893
Skewness	0.3277	0.1517	0.4191	0.6754	0.9306	4.9632	4.5803
Kurtosis	3.0401	2.8772	3.079	3.8952	4.6588	39.2973	38.4247
Min	-2.5208	-2.7928	-2.5555	-2.499	-2.3896	-1.6897	-1.6897
Max	3.6891	3.0305	3.9088	4.7459	5.384	15.8291	14.5264
T=4,000							
Mean	0.0125	0.0129	-0.0091	-0.0174	-0.032	-0.0047	0.0736
Median	-0.0284	0.0207	-0.0343	-0.0698	-0.101	-0.1906	-0.2388
Interquartile	1.4018	1.3824	1.3787	1.3663	1.3466	1.361	1.2773
Std	1.0037	1.0032	0.9764	1.0413	1.0338	1.1236	1.6586
Skewness	0.299	0.065	0.1761	0.4114	0.5107	2.5336	5.0509
Kurtosis	2.9486	2.7182	2.7838	3.1906	3.3683	26.6224	41.7284
Min	-2.5889	-2.6396	-2.9119	-2.6352	-2.6287	-2.1394	-1.6897
Max	3.7365	3.0624	2.9077	4.0682	4.0697	13.9347	15.6304
T=8,000							
Mean	0.0338	0.0119	-0.005	-0.0153	0.0005	-0.0085	0.0048
Median	-0.0158	-0.0069	-0.0183	-0.0475	-0.0628	-0.0953	-0.1623
Interquartile	1.3413	1.4163	1.3631	1.339	1.2612	1.3258	1.3398
Std	1.0084	1.0072	0.9816	1.0195	1.0204	1.0172	1.2082
Skewness	0.1865	0.0216	0.1598	0.2617	0.3875	0.6818	4.0959
Kurtosis	2.9969	2.7829	2.9478	3.3169	3.421	4.0203	54.2844
Min	-2.7498	-2.8376	-2.7588	-3.0098	-2.6839	-2.4151	-2.1616
Max	4.0113	3.0602	3.193	3.8564	3.8725	5.2102	18.1697
T=16,000							
Mean	0.01	-0.0391	-0.046	0.0023	0.0003	-0.0024	-0.0144
Median	0.021	-0.014	-0.0713	-0.0698	-0.0434	-0.0981	-0.1207
Interquartile	1.4854	1.3538	1.3109	1.3941	1.3128	1.2777	1.3705
Std	1.0372	0.9754	0.9765	1.0046	0.9596	0.9564	1.0077
Skewness	0.0616	-0.0375	0.1682	0.2513	0.2964	0.4948	0.6572
Kurtosis	2.8417	2.8552	2.9389	3.0173	3.1056	3.3612	3.8727
Min	-3.1526	-3.6499	-3.2333	-2.8064	-2.5062	-2.252	-2.4444
Max	3.6007	2.8835	4.0139	3.8634	3.6276	3.5028	4.7747
T=32,000							
Mean	0.0041	-0.0255	-0.0411	-0.0031	0.0106	-0.0002	0.0046
Median	0.0111	-0.032	-0.0407	-0.0249	-0.0362	-0.0425	-0.0242
Interquartile	1.3653	1.2966	1.3	1.2536	1.2963	1.3026	1.3943
Std	1.0259	0.9897	0.9845	0.9947	0.9829	0.9741	1.0101
Skewness	-0.077	-0.0456	0.0201	0.1984	0.1342	0.2338	0.2911
Kurtosis	3.0131	3.0823	2.9908	3.1692	3.5089	3.0939	3.1006
Min	-3.3324	-3.1044	-2.8953	-2.9944	-4.1167	-3.2703	-2.8103
Max	2.9448	3.45	3.2925	3.5756	3.5953	3.9835	4.3914
T=64,000							
Mean	-0.0099	-0.0268	-0.0411	-0.012	0.0295	0.0304	0.0245
Median	0.0052	-0.0681	-0.0356	-0.0206	0.0546	0.0376	-0.0126
Interquartile	1.3781	1.3325	1.3489	1.3604	1.3289	1.3956	1.4232
Std	1.0171	0.9958	0.9739	0.9954	1.0209	1.0108	1.0283
Skewness	0.0141	0.0541	0.0222	0.0795	0.0694	0.0275	0.1892
Kurtosis	2.9655	3.07	3.0996	3.0128	3.2001	2.9128	2.9231
Min	-3.6939	-3.4258	-3.1378	-3.3478	-3.4211	-3.0548	-2.5846
Max	3.2602	3.1786	3.224	3.0645	3.1615	3.3964	3.6168

Note: [1] Summary statistics are conditional on spending a positive number of time periods in State 2.

Figure 3.12: Distribution of the rescaled sampling error $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$ when states are observed[1].



Note: [1] The distribution is conditional on spending a positive number of time periods in State 2. The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000.

$T^{-\frac{1}{2}}$. There is also evidence to suggest that the Maximum Likelihood Estimator (MLE) of θ_1 is mean and median unbiased.

The rescaled sampling error in copula parameter θ_2 in state 2

Next let's consider the distribution of the standardised sampling error associated with *state 2*. Figure 3.12 shows the sampling distribution of θ_2^s in the low occupancy state. The behaviour along the main diagonal seems to be consistent with the heuristic theory. That is, the distribution of θ_2^s is not converging to the standard normal because the amount of information about the *state 2* is not going to infinity. No matter how long one makes the run length T, one will not spend an increasing number of time periods in *state 2*, because under the alternative asymptotic sequence Case 1, the number of time periods spent in *state 2* remains stochastically bounded away from infinity. Whilst it is theoretically conceivable that the number of time periods spent in *state 2* could be very large if one just happens to have a really unusual set of realisations, it would nevertheless be very unlikely. Hence, the sampling distribution of a parameter in *state 2* is not going to concentrate down. That is, the plots indicate that if matrix P is allowed to vary in the fashion described in (3.38), this will induce a constant expected number of time periods spent in *state 2*. Nevertheless, the rescaled distribution of the sampling error is not far from the standard normal. It can be noticed that, if one makes G matrix sufficiently large, then one will tend to be in *state 2* a lot more of the time. This characteristic would seem to indicate which diagonal one has to get to for the usual asymptotics to work reasonably well. If one moved sufficiently down to a lower finite diagonal, the parameter θ_2 would be reasonably well estimated most of the time, as a result of having more observations from *state 2*. This is consistent with the heuristic theory which would suggest that the larger the G matrix is, the further away the P matrix is from the boundary for a given run length. Consequently, the larger the G matrix, the greater the fraction of time one would spend in the low occupancy state, and the better things would behave. However, the usual asymptotics would still not apply because the amount of time spent in *state 2* would be stochastically bounded.

If one goes vertically down the column, both the run length T and matrix G will increase, whilst matrix P is being held constant. Hence, we would expect the usual asymp-

totics to work. In the usual asymptotic framework, the proportion of time spent in *state 2* is going to a constant, which means that the total number of time periods spent in *state 2* is increasing to infinity. Hence, the unconditional probability of being in *state 2* is non-zero.

We can also analyse how things behave when G and P change, whilst keeping the run length T constant. If one fixes T and varies P, then the number of time periods spent in *state 2* is getting smaller and smaller because P is getting closer and closer to the boundary case presented in (3.38). Therefore, one is simultaneously changing both the steady state probability of being in *state 2* and the probability of transiting from *state 2* to *state 2*. Consequently, this makes it less likely to ever be in *state 2*, and at the same time making it less likely to remain in *state 2* given one is currently in *state 2*. This results in the number of time periods spent in *state 2* to halve as one moves across the columns. The consequence of this is that there is less and less information about *state 2*, which in turn leads to poor estimates of the parameter θ_2. Therefore, the usual asymptotic theory would not work well in this setting.

On closer inspection of the summary statistics in Table 3.10, one can assess whether the degree of kurtosis or skewness is associated with different values of the matrix G. In accordance with our anticipations, it can be noted that the closer the matrix G is to zero, the more problematic is the behaviour of the rescaled sampling error. This can be seen in the top right part of the table where the median becomes strongly biased, whilst the mean is never strongly biased in most of the cells. In addition, the interquartile range increases, and so does the standard deviation. This region also tends to produce positive skewness and considerable excess kurtosis[8]. This would create issues for applying the usual asymptotic formulae of the covariance matrix. It is easier to discern considerable discrepancies by looking at the PP plots in Figure 3.28. The PP plots indicate that there are only noticeable discrepancies in the upper right corner. Below that region the PP plots are pretty close to a straight 45° line, with the black and the red line essentially being

[8] When there are no observations from *state 2*, whatever estimate of θ_2 is obtained, it will spurious. For this reason, realisations with no observations from *state 2* have been omitted from the estimation. There are six cells in total for which this is the case.

indistinguishable from one another. Also one can notice that it takes longer for the usual asymptotics to work reasonably well in the last column.

The rescaled sampling error in transition probability parameter p_{11}

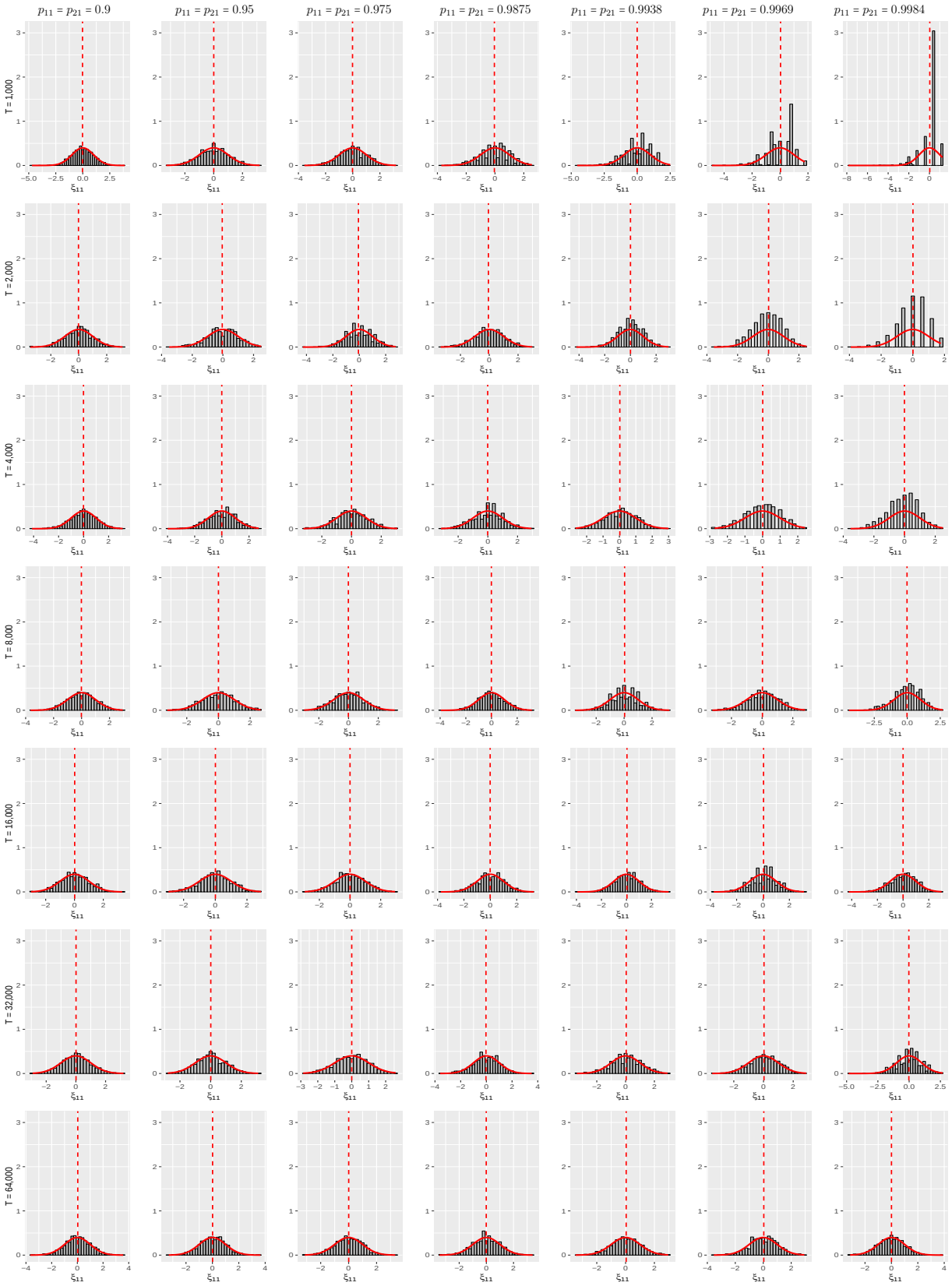
Now let's consider Figure 3.13 which displays distributions of the rescaled sampling error in p_{11}, which we denote as $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$. Along the main diagonal it can be observed that the usual asymptotic theory seems to work well most of the time. This behaviour is not surprising because this is what the heuristic theory would suggest. As the run length T doubles, the expected number of transitions from *state 1* to *state 1* more than doubles by a small amount. This is because when one moves down the main diagonal, one is simultaneously changing the probability of transiting from *state 1* to *state 1*, so that it is more likely that one remains in *state 1* given one is currently in *state 1*. As a result, most of the time is spent in high occupancy *state 1*. Consequently, the information about the parameter p_{11} is increasing to infinity and, therefore, it is consistently estimated. In addition, there do not seem to be considerable issues, such as skewness, arising as a result of transition probability matrix P getting closer to the boundary case.

If one fixes T and varies P and G, then the expected number of transitions from *state 1* to *state 1* increases by slightly less than g_{11}. This is the behaviour observed as one moves across columns. As a result of G getting close to zero, an additional information about the transition parameter p_{11} is obtained, and hence, one would expect that the sampling behaviour would improve only so much. However, the closer the G matrix is to zero, the closer is the transition probability matrix P to the boundary. An additional information obtained is not sufficient to reduce the uncertainty about the estimate of p_{11} in line with the advancement towards the boundary of the parameter space. Consequently, the distribution becomes deeply skewed. Furthermore, the number of transitions from *state 1* to *state 2*, approaches zero. Due to the properties of MLE of the transition probabilities of a Markov chain, this will result in obtaining an estimate of the transition probability p_{11} as being equal to one with increasingly higher probability. This intrinsic characteristic results in a plot that exhibits a sharp peak at the positive value of $(1 - p_{11})$. Hence, it is important to note that the abundance of estimates being equal to one is not due to the numerical

Table 3.11: Summary statistics for the rescaled sampling error $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$ when states are observed.

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	-0.0489	-0.0535	-0.0641	-0.0604	-0.0193	0.0189	0.0113
Median	-0.0558	-0.0669	0.0693	0.0967	0.083	0.0639	0.4482
Interquartile	1.3226	1.4291	1.4676	1.451	1.6202	1.1388	0.8028
Std	1.0133	1.0369	1.0186	1.0142	1.0068	0.9835	0.9588
Skewness	-0.1892	-0.1606	-0.149	-0.2861	-0.5135	-0.5372	-0.7523
Kurtosis	3.4539	2.8071	2.9931	2.8936	3.342	3.2205	3.4858
Min	-4.7001	-3.2901	-3.5577	-3.425	-4.4154	-3.95	-3.582
Max	3.3173	2.9831	3.1784	2.3973	2.4987	1.7669	1.2494
T=2,000							
Mean	-0.0196	-0.042	-0.0618	-0.0553	0.0045	0.0231	0.0103
Median	0.0079	-0.0079	-0.0435	-0.0658	0.1192	0.0915	0.0673
Interquartile	1.2864	1.3429	1.4857	1.2309	1.4324	1.6101	1.1351
Std	0.9974	1.015	1.0406	0.9909	0.9931	0.9858	0.9687
Skewness	-0.1638	-0.2184	-0.1321	-0.2121	-0.2731	-0.305	-0.5405
Kurtosis	3.2607	2.9439	3.0716	3.1384	3.2892	2.9696	3.2401
Min	-3.4643	-3.5039	-4.2769	-3.375	-4.2086	-3.5475	-3.9194
Max	3.1325	2.4994	2.7588	2.9882	2.9683	2.4994	1.7673
T=4,000							
Mean	-0.0103	-0.0048	-0.0016	-0.012	0.0238	0.0091	-0.0104
Median	0.0111	0.0689	0.0436	0.0519	-0.0327	0.1303	0.0958
Interquartile	1.3562	1.4644	1.4667	1.3051	1.4175	1.4233	1.6048
Std	1.0395	1.0468	1.0358	1.0107	0.9848	0.9885	0.9983
Skewness	-0.1912	-0.232	-0.1316	-0.1706	-0.0976	-0.2069	-0.3014
Kurtosis	3.1878	2.9957	2.928	2.9517	2.7612	2.7147	2.8874
Min	-4.1198	-3.9883	-3.2364	-3.1587	-2.6764	-2.7227	-3.5237
Max	3.1183	2.719	3.0785	2.9385	2.9941	2.4024	2.4997
T=8,000							
Mean	-0.0122	-0.0004	-0.0124	-0.0101	0.0344	0.0215	-0.0178
Median	0.0138	0.05	-0.0009	-0.0266	0.097	-0.0163	0.1359
Interquartile	1.3897	1.4199	1.3402	1.3369	1.2888	1.4087	1.4187
Std	1.0208	1.0374	0.996	1.0196	1.0008	1.0169	1.0268
Skewness	-0.0649	-0.0944	-0.0156	-0.0654	-0.1083	-0.2083	-0.419
Kurtosis	2.8404	2.8374	2.72	3.0367	3.1416	2.9559	3.1894
Min	-3.6513	-3.2297	-2.9908	-3.9455	-3.4929	-3.2431	-4.4126
Max	2.9456	2.6562	3.1066	3.1398	3.2391	2.7965	2.6864
T=16,000							
Mean	0.0131	0.0075	-0.0011	-0.0139	0.0392	0.0387	-0.0007
Median	0.0097	0.0177	-0.0038	-0.0358	0.0377	0.1193	-0.0081
Interquartile	1.3066	1.2843	1.3409	1.3425	1.2157	1.2807	1.4043
Std	0.9741	0.9958	0.9982	1.026	1.0098	0.9866	1.026
Skewness	0.0327	-0.0683	-0.0705	-0.171	-0.0237	-0.1893	-0.3743
Kurtosis	2.8894	3.0443	2.8437	2.9635	3.1254	3.1579	3.1551
Min	-3.0158	-2.9555	-2.96	-3.7487	-3.9208	-3.7295	-4.227
Max	3.3105	2.8892	2.9341	3.2889	3.3721	3.104	2.9985
T=32,000							
Mean	0.0089	0.0094	0.0002	-0.0316	0.0411	0.0333	-0.021
Median	-0.0108	0	0.0349	-0.0025	0.0529	0.0687	-0.0113
Interquartile	1.2435	1.3091	1.3226	1.3368	1.3617	1.4086	1.4187
Std	0.9537	0.9724	0.9712	1.0332	1.0509	1.008	1.013
Skewness	0.1083	0.0549	-0.0695	-0.2352	-0.1339	-0.2529	-0.2916
Kurtosis	3.039	2.9263	2.6938	3.2001	2.9577	2.9287	3.3573
Min	-2.9159	-2.7948	-2.867	-3.598	-3.3879	-3.4571	-4.8399
Max	3.2082	3.1047	2.5885	3.5317	2.914	2.8839	2.6826
T=64,000							
Mean	0.0297	0.0151	0.0049	-0.0094	0.0396	0.0386	0.0039
Median	0.0264	0.0312	0.0256	-0.036	0.0247	0.0266	-0.0158
Interquartile	1.2434	1.3249	1.3373	1.3778	1.3669	1.3519	1.3041
Std	0.9645	0.9704	0.9985	1.017	1.0212	0.9779	1.0023
Skewness	-0.0331	-0.036	-0.098	-0.0698	-0.0525	-0.0808	-0.0777
Kurtosis	3.2684	3.2582	3.109	3.2913	2.9911	3.06	3.1275
Min	-3.6638	-3.3034	-3.1363	-3.4484	-3.4727	-3.5341	-3.1273
Max	3.5654	3.6404	3.3322	3.4017	2.9584	2.942	3.894

Figure 3.13: Distribution of the rescaled sampling error $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$ when states are observed.



Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000

rounding error issues associated with being close to the boundary of the parameter space. Summary statistics displayed in Table 3.11 reveal that, when matrix G is very close to zero, and the sample size is close to $T = 1000$, the MLE of p_{11} is strongly median biased. However, the mean is never strongly biased.

The usual asymptotic sequence is described as we move vertically down the column. This implies that the transition probability matrix P is being held fixed, whilst matrix G and the run length T are allowed to vary. This results in the expected number of transitions to double as the run length T doubles. The consequence of this is that there is more and more information about the transitions from *state 1* to *state 1*, so that the parameter p_{11} is well estimated, and therefore the usual asymptotics work well.

The rescaled sampling error in transition probability parameter p_{21}

Next we consider the behaviour of the rescaled sampling error in the transition parameter from *state 2* to *state 1*, denoted as ξ_{21}. In Figure 3.14 the main diagonal displays the sampling behaviour under the alternative asymptotic sequence Case 1. As the run length T doubles, the expected number of transitions from *state 2* to *state 1* is slowly increasing to $g_{11} = 100$. This happens as a result of a simultaneous change in the probability transition matrix P towards the boundary case, so that it is less likely that one remains in *state 2* given one is currently in *state 2*. Therefore, the expected number of transitions from *state 2* to *state 2* is decreasing. Subsequently, this will result in estimates of p_{21} being equal to one with increasingly higher probability. This explains the multitude of parameter estimates being equal to one, which is not a result of the numerical rounding error issues. It is evident from Figure 3.14 that under the alternative asymptotic sequence Case 1, the limiting distribution is not the standard normal distribution. Furthermore, it seems like the deviation from the standard normal is not stable, and indisputably intensifying as the run length T gets larger. Therefore, one would expect the supremum of the difference between the standardised sampling error and the standard normal distribution to be increasing.

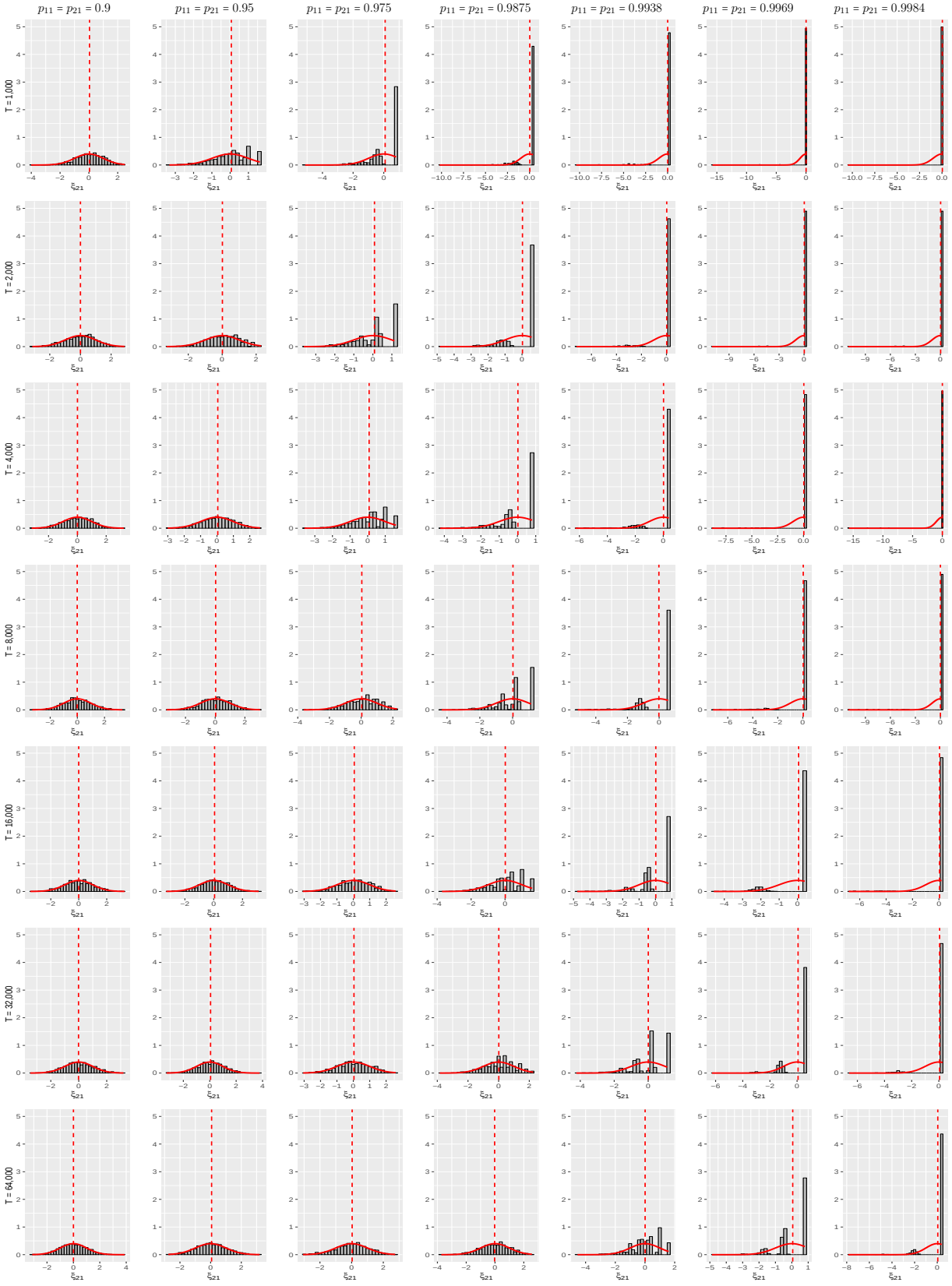
From the plots one can learn that there is a variety of ways in which the regular asymptotic theory would be wrong. First, there is higher and higher probability that the estimates of p_{21} would lie on the boundary of the parameter space. Hence, as one moves

Table 3.12: Summary statistics for the rescaled sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are observed.[1]

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	0.0395	0.0612	0.0911	0.0314	0.0093	0.0192	0.0363
Median	0.0927	0.1189	0.8002	0.3976	0.1982	0.0989	0.0494
Interquartile	1.3154	1.4077	1.2804	0	0	0	0
Std	0.9901	0.9603	0.9388	0.994	0.9574	0.8111	0.3719
Skewness	-0.381	-0.4457	-1.2554	-3.3979	-5.7710	-12.5021	-28.3020
Kurtosis	3.2945	2.993	4.3947	19.8273	40.5249	189.7384	802.0012
Min	-4.0041	-3.3253	-5.1275	-10.2045	-10.3706	-15.7293	-10.4945
Max	2.4987	1.6214	0.8002	0.3976	0.1982	0.0989	0.0494
T=2,000							
Mean	0.0006	0.0121	0.0854	0.0254	0.0196	0.0229	0.0260
Median	0.0807	0.0704	0.2079	0.5624	0.2803	0.1399	0.1399
Interquartile	1.3395	1.3989	1.7087	1.3434	0	0	0
Std	0.9896	0.9608	0.9373	0.9668	0.9699	0.8767	0.8629
Skewness	-0.2531	-0.3039	-0.7084	-1.6953	-4.0017	-8.5381	-8.7130
Kurtosis	2.8246	2.9629	3.0298	5.3119	19.8805	83.9897	88.1432
Min	-3.1419	-3.2796	-3.6343	-4.8367	-7.1948	-11.0551	-11.0551
Max	2.8801	2.2936	1.132	0.5624	0.2803	0.1399	0.1399
T=4,000							
Mean	0.0058	0.0474	0.0835	0.0242	0.0199	0.0335	0.0084
Median	0.0529	0.0791	0.2088	0.7955	0.3965	0.1979	0.0989
Interquartile	1.4375	1.3559	1.3846	1.3258	0	0	0
Std	1.0267	0.9773	0.9498	0.9788	0.9969	0.9386	1.1010
Skewness	-0.1887	-0.21	-0.557	-1.1163	-2.7644	-6.1602	-12.5458
Kurtosis	2.8875	2.787	3.2038	3.6722	10.7684	42.8437	163.4872
Min	-3.1742	-3.0059	-3.5916	-4.2285	-6.281	-8.8502	-15.7229
Max	3.2242	2.6261	1.6011	0.7955	0.3965	0.1979	0.0989
T=8,000							
Mean	-0.0123	0.0295	0.042	0.0226	-0.0154	0.0344	0.0013
Median	-0.0465	0.0646	0.1282	0.1972	0.5607	0.2799	0.1399
Interquartile	1.3893	1.3716	1.4685	1.7476	1.5206	0	0
Std	1.0426	0.9714	0.9655	1.0039	1.0129	0.9588	1.0232
Skewness	-0.0193	-0.1181	-0.3732	-0.8084	-1.7547	-4.0113	-7.9433
Kurtosis	2.9307	2.9513	2.9062	3.3373	5.8999	19.0226	68.8264
Min	-3.3247	-3.2772	-3.6231	-4.3296	-5.1356	-7.1848	-11.0485
Max	3.309	3.0252	2.2644	1.125	0.5607	0.2799	0.1399
T=16,000							
Mean	-0.0041	0.0589	0.0557	0.0138	0.0181	0.0624	0.0433
Median	-0.0082	0.0563	0.1141	0.2405	0.793	0.3959	0.1978
Interquartile	1.4091	1.3288	1.3831	1.4725	1.3081	0	0
Std	1.0371	0.9885	0.9865	0.9999	0.9997	0.893	0.8683
Skewness	-0.0268	-0.1101	-0.3299	-0.5301	-1.345	-2.4659	-5.6902
Kurtosis	2.9009	3.1172	2.9535	3.0224	4.8777	7.7684	34.8855
Min	-3.4651	-3.2932	-3.0866	-3.7125	-4.7721	-4.5722	-6.8348
Max	3.2752	3.1223	2.5821	1.5911	0.793	0.3959	0.1978
T=32,000							
Mean	-0.0163	0.062	0.0386	0.0106	0.0071	0.0727	0.0545
Median	-0.0578	0.1096	0.0403	0.0441	0.2152	0.5599	0.2797
Interquartile	1.4112	1.331	1.4412	1.3863	1.7855	0	0
Std	1.0516	1.032	1.0147	1.0019	1.0086	0.9541	0.8908
Skewness	0.0637	-0.132	-0.1473	-0.3651	-0.9255	-2.1016	-3.9797
Kurtosis	2.8272	3.0645	2.6865	3.0881	3.8017	8.1438	18.4977
Min	-3.4657	-3.2718	-3.0589	-3.8372	-4.235	-6.2654	-6.6058
Max	3.197	3.775	2.5784	2.2501	1.1215	0.5599	0.2797
T=64,000							
Mean	0.003	0.0616	0.0321	-0.026	0.0082	0.0546	0.0452
Median	-0.0125	0.0525	0.0642	0.0217	0.238	0.7918	0.3956
Interquartile	1.3772	1.3213	1.3516	1.3579	1.4084	1.2927	0
Std	1.0051	1.0017	1.0295	1.0204	1.0082	0.945	0.973
Skewness	0.0604	0.0562	-0.2278	-0.281	-0.6596	-1.1678	-3.0001
Kurtosis	3.0186	2.9654	2.9196	2.9713	3.5757	4.0835	13.2121
Min	-3.1155	-2.7626	-3.0658	-3.7262	-4.5437	-4.8702	-7.8602
Max	3.8844	3.1942	2.7498	2.562	1.5861	0.7918	0.3956

Note: [1] Summary statistics are computed conditional on spending a positive number of time periods in State 2.

Figure 3.14: Distribution of the rescaled sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are observed.[1]



Note: [1] The distribution is conditional on spending a positive number of time periods in State 2. The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000.

further down the main diagonal in Figure 3.14, it can be observed that the usual asymptotic theory would allocate a substantial positive probability mass outside the boundary, that is the non-feasible region beyond the value of $(1 - p_{21})$. This would result in the distortion away from the standard normal distribution by showing considerable deep skewness. This is not surprising since the standard normal distribution does not provide an adequate approximation when the mean of a distribution lies close to the boundary of the parameter space. Second, the distribution of the rescaled sampling error noticeably accumulates at the positive value of $(1 - p_{21})$. In addition, one can notice strong skewness characterised by large negative values outside the region where the standard normal distribution would assign most of its probability mass. Third, plots reveal that there is not enough probability mass in the centre of a distribution, which is conceivably a more serious distortion from normality. One can also notice that as both p_{21} and p_{11} approach the value of 1, the behaviour of their rescaled sampling errors, ξ_{21} and ξ_{11} respectively, considerably differs from each other. Indeed, the heuristic theory would suggest that the limiting distribution of ξ_{21} is a degenerate distribution with all of its probability mass centred at the $\lim_{p_{21} \nearrow 1}(1 - p_{21})$. This can be deduced by noting that the expected number of transitions from *state 2* to *state 2* along the main diagonal is approaching $\lim_{T \to \infty} \mathbb{E}\left(n_{22}^{T}\right) = 0$, as it is shown in (3.45). According to the analytic formulae of the MLE, the $\widehat{p}_{21}$ is computed as the number of times *state 2* is followed by *state 1*, divided by the total transitions from *state 2* to any state. Therefore, regardless of how many times one visits *state 2*, if it is not followed immediately by *state 2*, the estimate of p_{21} will always equal 1 by construction of the MLE. This will result in the MLE of the transition probability p_{21} to be biased in finite samples, although being consistent.

Under the usual asymptotic sequence, that is as we move vertically down the column, the expected number of transitions from *state* 2 to *state* 1 doubles as the run length T doubles. The consequence of this is that there is more and more information about the transitions from *state* 2 to *state* 1, so that the parameter p_{21} is well estimated. Therefore, the behaviour of the rescaled sampling error ξ_{21} approaches that of a standard normal. Also, the expected number of transitions from *state* 2 to *state* 1 doubles as the run length

T doubles.

If one moves along any row across columns, one changes matrices P and G whilst holding run length T constant. This induces a simultaneous change in the steady state probability of being in *state* 2, and the probability of transiting from *state* 2 to *state* 1. Consequently, the process is less likely to be in *state* 2. Furthermore, given the process is already in *state* 2, it is less likely to remain in *state* 2. This results in the expected number of transitions to decrease by slightly less than a half as one moves across the columns. Therefore, the information about the transition parameter p_{21} decreases, and the general asymptotic theory does not apply. The closer the G matrix is to zero for a given run length, the worse things behave.

In summing up the simulation results, there seem to be three interconnected factors at work that drive the observed distortions under the alternative asymptotic sequence Case 1. First, is that the expected number of time periods spent in *state 2* is stochastically bounded. That is, no mater how large one makes the run length T by moving very far down the alternative asymptotic sequence, the expected number of time periods spent in *state 2* is unlikely to increase. In other words, the probability of seeing the number of time periods spent in *state 2* being really large is very small. Second, this alternative asymptotic sequence pushes the parameter p_{21} to the upper boundary, which creates distortions. When one is near the boundary, normal distribution may not work necessarily well, because it might be assigning a substantial probability mass outside the parameter space. Third, due to the transition probability parameters being close to the boundary of the parameter space, the distribution of the rescaled sampling error results in having a sharp peak and inherent discreteness. This is due to the way the MLE of p_{21} is constructed.

Unobserved states

Let's now consider a more plausible world in which one cannot observe the states. When states cannot be observed, it adds an additional aspect to the estimation problem. This undoubtedly makes it interesting to assess what are the consequences of not being able to observe the states. The empirical results have been obtained for both observed and unobserved case, so it is possible to carry out comparison on what is the nature and the

magnitude of the differences in these results, and what seems to be driving these differences. The simulation results in the observed case confirmed that there is inherently going to be a problem even if one could observe the states, because the number of time periods spent in *state 2* is not increasing to infinity. Hence, the estimates of the copula parameter for that state will not be converging to the true parameter value, but will remain blurred and spread out.

The rescaled sampling error in copula parameter θ_1 in state 1

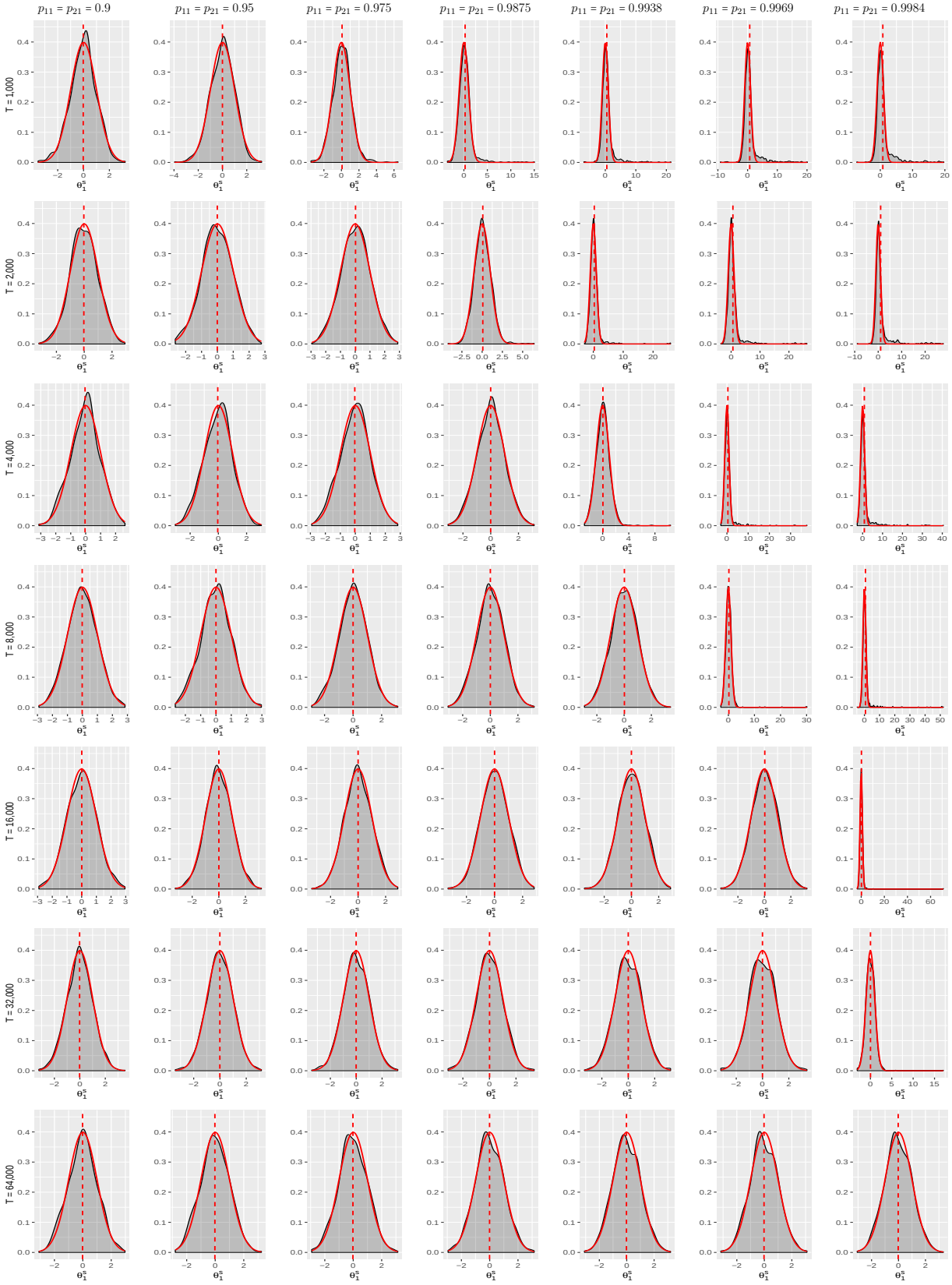
In Figure 3.15 along the main diagonal we can notice that the behaviour of the rescaled sampling error under the alternative asymptotic sequence is not substantially affected by the lack of observability of the states. The behaviour under the usual asymptotics seems also not to be considerably affected. This is not surprising because one gets a lot of information about the copula parameter θ_1 in *state 1*, and one is not getting close to the boundary point of a parameter space, because the transition probability matrix P is being held fixed.

However, the behaviour of the rescaled sampling error markedly deteriorates as G gets very close to zero. This can be observed by moving along any row across the columns, which imposes the run length T to be fixed whilst changing matrices P and G. The summary statistics in Table 3.13 indicate strong mean bias and relatively milder median bias of the MLE. We can also detect a substantial increase in uncertainty, deeper skewness and stronger kurtosis. These distortions happen in the upper right region of Figure 3.15 and the corresponding region in Table 3.13. In these regions, the number of transitions from *state 1* to *state 2* is very low, and so is the expected number of time periods spent in *state 2*. In the observed case it was noticed that as a result of spending a negligible number of time periods in *state 2*, it translated into distortions in the distribution of the rescaled sampling error in p_{11}. Although the likelihood function for the unobserved case is different, it is expected that a similar nature of distortions would be present in this case, such as severe kurtosis, considerable skewness, large negative values, and high uncertainty in the parameter estimates. It is through the construction of the likelihood function that these distortions, associated with low occupancy *state 2*, "contaminate" the estimates of

Table 3.13: Summary statistics for the rescaled sampling error $\theta_1^s = (\hat{\theta}_1 - \theta_1)/se(\hat{\theta}_1)$ when states are not observed.

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	-0.0241	-0.0045	0.0712	0.258	0.4989	0.7226	0.7859
Median	0.032	0.0364	0.0674	0.1207	0.1843	0.2569	0.2719
Interquartile	1.2847	1.3307	1.315	1.3704	1.3927	1.4405	1.4522
Std	1.0013	0.9874	1.0736	1.5221	2.1434	2.5586	2.5842
Skewness	-0.1906	-0.1829	0.3345	3.512	4.2634	3.4162	3.6936
Kurtosis	3.2541	3.2043	4.8828	27.967	31.9466	20.3767	21.5365
Min	-3.5181	-3.9573	-3.4735	-3.5053	-7.0123	-9.1715	-7.2812
Max	3.2177	3.2241	6.4576	15.0475	21.4938	20.1206	19.6761
T=2,000							
Mean	-0.0266	-0.0077	0.0211	0.0794	0.3136	0.619	1.0469
Median	-0.0307	-0.0384	0.037	0.0244	0.1133	0.152	0.2348
Interquartile	1.3132	1.3328	1.3247	1.3449	1.3087	1.3892	1.492
Std	0.9855	0.977	0.9891	1.0512	1.9575	2.6823	3.6915
Skewness	-0.051	-0.0436	-0.0179	0.5344	7.5599	5.6097	4.4205
Kurtosis	2.9463	2.7839	2.8124	5.5442	90.1374	44.4286	25.7951
Min	-3.2589	-2.6567	-2.9553	-4.2076	-2.9608	-3.5546	-8.9242
Max	2.9323	2.7739	2.8884	6.4213	25.8446	26.2346	27.7361
T=4,000							
Mean	-0.0412	-0.0436	-0.0355	-0.0042	0.0417	0.455	0.8524
Median	0.0283	0.023	0.0065	0.017	0.0454	0.1471	0.1967
Interquartile	1.3078	1.3413	1.3147	1.3477	1.327	1.4137	1.4936
Std	0.9825	0.9898	0.9872	0.9977	1.0965	2.6837	3.6897
Skewness	-0.1111	-0.1256	-0.0971	-0.0188	1.3205	7.7765	6.1287
Kurtosis	2.8202	2.7877	2.802	2.9645	12.8807	87.1306	49.9224
Min	-3.1256	-3.0527	-2.9126	-3.0722	-2.783	-2.8627	-2.7084
Max	2.6277	3.0458	2.8314	3.1283	10.3888	37.7109	40.5088
T=8,000							
Mean	-0.016	-0.0012	-0.0013	0.003	0.0201	0.1954	0.6511
Median	-0.0337	0.0291	0.0209	-0.0054	0.0559	0.0621	0.0883
Interquartile	1.3276	1.2767	1.3163	1.3393	1.3196	1.351	1.4028
Std	0.9889	0.9889	0.9869	0.9876	0.9871	1.8909	4.1106
Skewness	0.0247	-0.0113	-0.066	-0.057	-0.0214	9.685	9.0691
Kurtosis	2.8186	2.8929	2.9571	2.9456	2.9026	141.2008	99.6973
Min	-2.9032	-2.6584	-2.9015	-3.0586	-2.9106	-2.8852	-4.737
Max	2.8375	2.9666	3.1302	3.175	3.4079	30.129	52.0006
T=16,000							
Mean	0.0075	0.0037	0.0008	0.0043	-0.0035	0.0297	0.1956
Median	0.0082	-0.0288	-0.017	-0.0015	0.0052	0.0351	0.0806
Interquartile	1.3676	1.3271	1.3288	1.3728	1.3452	1.3568	1.3301
Std	1.0058	0.9888	0.9893	0.9779	0.9893	0.992	2.5086
Skewness	-0.0028	0.004	-0.0128	-0.0622	-0.0929	-0.039	22.6185
Kurtosis	2.9268	3.0195	2.8919	2.8729	2.9222	3.0206	636.8162
Min	-2.9103	-3.2652	-3.3712	-3.3039	-3.4503	-3.2897	-3.4406
Max	2.9491	3.1936	2.8813	2.821	2.8693	3.2152	71.0131
T=32,000							
Mean	-0.0177	-0.0096	-0.0064	0.009	-0.0027	-0.003	0.0195
Median	-0.0155	-0.016	-0.03	0.0001	-0.0467	-0.007	-0.019
Interquartile	1.3377	1.3282	1.3594	1.3116	1.3785	1.4079	1.3922
Std	1.014	1.0018	1.0125	1.0103	1.022	1.0118	1.1555
Skewness	-0.0334	-0.0541	-0.0187	0.0153	0.0283	0.0216	3.2978
Kurtosis	3.038	3.0789	3.0587	3.0796	2.9973	3.1216	49.7778
Min	-3.2156	-3.4001	-3.4444	-3.1129	-3.3167	-3.1708	-3.0798
Max	3.5473	3.1508	3.2132	3.3656	3.1985	3.3784	17.0563
T=64,000							
Mean	-0.0178	-0.0223	-0.0218	-0.0088	-0.019	-0.0223	-0.0237
Median	-0.0102	-0.024	-0.0453	-0.0466	-0.0516	-0.0716	-0.0548
Interquartile	1.335	1.3789	1.3544	1.3646	1.4106	1.3522	1.3759
Std	1.0202	1.0089	1.0099	1.0048	1.0048	1.0047	0.9995
Skewness	-0.0143	0.0067	0.0089	0.0551	0.0622	0.029	0.0456
Kurtosis	2.8006	2.7604	2.9033	2.9415	2.8729	2.9045	2.8701
Min	-3.0918	-2.7973	-2.9227	-2.9427	-3.0869	-3.1624	-3.0629
Max	2.9693	3.2347	3.1313	3.1433	3.2013	3.098	3.3231

Figure 3.15: Distribution of the rescaled sampling error $\theta_1^s = (\hat{\theta}_1 - \theta_1)/se(\hat{\theta}_1)$ when states are not observed.



θ_1. However, unlike in the observed case, the abundance of the parameter estimates being equal to one may well be due to the numerical rounding error issues. When states cannot be observed, the probability of being in a particular state has to be estimated from the data given the copula parameters in both states. These results indicate that when the expected number of time periods spent in both states is substantial, then the lack of observability would not be to a great extent important. However, if the expected occupancy time in a particular state is negligible, then the distortions in the parameter estimates associated with the low occupancy state will "contaminate" the parameter estimates associated with the high occupancy state. Therefore, the lack of observability would have a stronger impact on the distribution of the rescaled sampling error.

The rescaled sampling error in copula parameter θ_2 in state 2

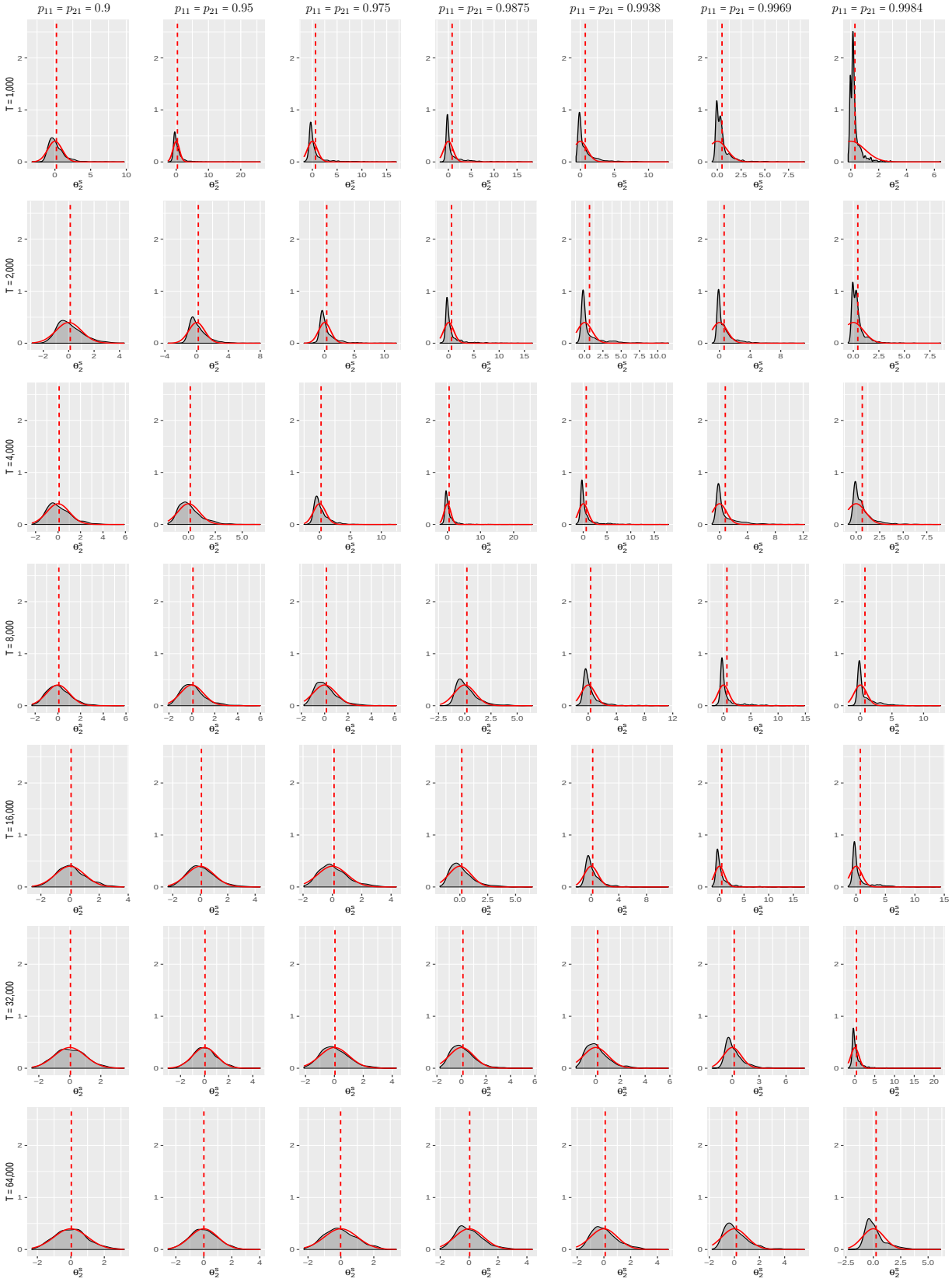
Now we examine the consequences of not being able to observe the states on the behaviour of the rescaled sampling error θ_2^s. In Figure 3.16 along the main diagonal we can notice a sizeable distortion in the behaviour of the rescaled sampling error under the alternative asymptotic sequence. When conducting a visual inspection of the plots, the presence of a deep positive skewness can be easily noticed, accompanied by high kurtosis and strong mean bias. Table 3.14 presents summary statistics confirming these distortions. We can easily calculate the expected number of time periods spent in *state 2* to be 100 using (3.41), and it would seem that this amount of information would not be sufficient for the usual asymptotics to work well when the states cannot be observed. The simulation results from the observed case suggested that if states could be observed, this amount of information would be sufficient for the usual asymptotics to describe reasonably well the behaviour of the rescaled sampling error. Therefore, we could obtain more information if we moved to a lower sub-diagonal, where the expected number of time periods spent in *state 2* would increase, and we would expect the behaviour to improve. Indeed, we can notice that the behaviour under the alternative asymptotics is reasonably well approximated by the standard normal distribution as we move down to lower sub-diagonals.

The behaviour under the usual asymptotics seems to be unsurprisingly unaffected. Although, as it was mentioned previously, one would need a somewhat larger sample for

Table 3.14: Summary statistics for the rescaled sampling error $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$ when states are not observed.

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	0.2142	0.3838	0.6396	0.8263	0.7307	0.4903	0.297
Median	-0.0481	-0.0781	-0.0922	-0.0602	0.0629	0.2705	0.1712
Interquartile	1.3386	1.2916	1.0628	1.1907	1.1379	0.6452	0.266
Std	1.2057	1.7274	2.0994	2.1514	1.4857	0.8498	0.4672
Skewness	2.0834	6.0144	3.4425	3.3227	2.7588	3.4336	4.3646
Kurtosis	11.75	69.2473	18.556	18.0814	13.3652	21.9872	38.963
Min	-3.2258	-2.5567	-1.6852	-1.8195	-0.6203	-0.5519	-0.1868
Max	9.7197	26.0762	16.975	18.3819	12.9378	9.1917	6.4504
T=2,000							
Mean	0.1243	0.2058	0.3512	0.5602	0.6661	0.6107	0.4382
Median	-0.0637	-0.1	-0.1133	-0.11	-0.0758	0.0825	0.2556
Interquartile	1.3388	1.3331	1.2104	0.9573	1.0933	1.0018	0.6071
Std	1.0352	1.2013	1.527	1.8962	1.7254	1.2417	0.7313
Skewness	0.838	1.7448	2.976	3.454	2.7896	3.1115	3.6029
Kurtosis	3.828	8.4357	16.6546	18.5489	12.5581	17.6279	26.1744
Min	-2.8977	-3.618	-3.4573	-1.7193	-1.1702	-0.9564	-0.5044
Max	4.379	8.0175	12.0208	16.6062	11.4745	11.1202	8.5492
T=4,000							
Mean	0.0966	0.1617	0.2535	0.515	0.6295	0.8299	0.6535
Median	-0.0812	-0.0822	-0.1463	-0.0929	-0.0769	0.0355	0.3455
Interquartile	1.3906	1.3634	1.2759	1.1435	1.0332	1.4271	1.0531
Std	1.0879	1.1486	1.3619	2.1512	1.9929	1.81	1.1292
Skewness	0.9	1.4104	2.9279	5.3001	3.441	2.5718	2.5969
Kurtosis	4.253	6.3469	19.4398	44.1399	18.8309	11.0901	12.6662
Min	-2.3324	-1.9237	-2.517	-2.2342	-1.5101	-1.0567	-0.8431
Max	5.9044	6.7001	12.4565	25.7057	17.9748	12.2014	8.9751
T=8,000							
Mean	0.0835	0.0989	0.1518	0.2028	0.3449	0.6364	0.7246
Median	-0.0211	-0.0307	-0.0458	-0.0534	-0.0841	-0.0934	-0.0156
Interquartile	1.4159	1.2804	1.2318	1.241	1.0358	0.9613	1.3257
Std	1.0677	1.0624	1.0554	1.0823	1.425	1.9822	1.6121
Skewness	0.7455	1.0267	1.4105	1.5743	3.1752	3.0593	2.4543
Kurtosis	4.2745	4.8658	6.3291	6.7328	16.9411	14.3254	11.1297
Min	-2.2924	-2.0897	-1.7707	-2.3587	-1.7232	-2.0709	-1.9292
Max	5.8558	5.9862	6.1544	6.4604	11.4091	14.8505	12.7056
T=16,000							
Mean	0.0743	0.0832	0.1131	0.1706	0.2222	0.4776	0.7347
Median	-0.0001	-0.0297	-0.0434	-0.0651	-0.121	-0.1039	-0.1009
Interquartile	1.3107	1.3216	1.2931	1.274	1.1704	1.0968	1.108
Std	1.0146	1.0171	1.0097	1.0873	1.1759	1.8849	2.0153
Skewness	0.4225	0.638	0.7945	1.3871	2.3251	4.0823	2.7455
Kurtosis	3.2145	3.6653	3.7475	5.9607	13.5275	25.3694	12.3638
Min	-2.6192	-2.3545	-1.9371	-1.7823	-2.1778	-1.4766	-1.33
Max	3.7333	4.3826	4.3659	6.5199	11.1859	17.2948	14.3835
T=32,000							
Mean	0.0415	0.0342	0.0413	0.1215	0.1591	0.2185	0.3925
Median	-0.0202	-0.007	-0.0448	-0.0326	-0.0394	-0.0887	-0.0837
Interquartile	1.4174	1.2733	1.291	1.2417	1.2042	1.1555	1.0021
Std	1.0071	1.0037	0.9841	0.9869	1.0392	1.0936	1.5845
Skewness	0.2478	0.3708	0.6329	1.088	1.5962	1.9346	4.6666
Kurtosis	2.881	3.6145	3.6956	5.1893	7.2497	9.2389	43.1191
Min	-2.4478	-3.0454	-2.1069	-1.7501	-1.5925	-2.2456	-1.6921
Max	3.5193	4.6302	4.2963	5.7998	5.8959	8.0756	21.6642
T=64,000							
Mean	0.0282	0.0052	-0.0184	0.0373	0.0967	0.1592	0.2556
Median	-0.0141	-0.034	-0.0891	-0.0923	-0.0474	-0.0853	-0.0267
Interquartile	1.2701	1.3476	1.3597	1.2587	1.2494	1.2072	1.0521
Std	0.9789	0.991	0.971	0.9923	0.9984	1.032	1.1443
Skewness	0.3018	0.2657	0.3575	0.912	0.933	1.4491	1.9576
Kurtosis	2.9451	3.0108	2.8361	4.3087	4.0773	6.0911	8.2316
Min	-2.3291	-2.567	-2.2457	-2.0137	-2.0798	-1.7719	-2.2903
Max	3.1993	4.0498	3.3773	4.3861	4.8324	5.594	6.144

Figure 3.16: Distribution of the rescaled sampling error $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$ when states are not observed.



the usual asymptotics to start to work reasonably well.

Next, if we fix the run length T and move across the columns, the matrix G does not need to get very close to zero for the behaviour of the rescaled sampling error to start to deteriorate. Examination of the first cell for $T = 1000$ in Table 3.14 indicates the presence of distortions in the form of a strong positive skewness, high kurtosis and a sizeable mean bias equal to *0.21*. This can also be noticed by the visual inspection of the first plot in Figure 3.16.

The rescaled sampling error in transition probability parameter p_{11}

Now we consider Figure 3.17 which displays distributions of the rescaled sampling error in p_{11} when states are unobserved. Along the main diagonal the sampling behaviour under the alternative asymptotic sequence Case 1 is displayed. There is a noticeable impact as a result of not being able to observe the states. The most prominent is a moderate negative skewness and increasing kurtosis. Table 3.15 provides the summary statistics which confirm the observed distortions. It is interesting to note that the interquartile range is never far off from the reference distribution. It is not trivial to establish what the limiting distribution of the rescaled sampling error will be, however it seems plausible to conjecture that the limiting distribution will not be the standard normal distribution. As expected, the behaviour under the usual asymptotics is unaffected, although larger samples are required for distortions to wash away, with skewness being the most pronounced and the most persistent amongst others.

If one moves along any row by fixing T and changing matrices P and G, the behaviour of the rescaled sampling error deteriorates rapidly well before G gets very close to zero. These distortions appear in the form of deep negative skewness, severe kurtosis, very large negative values, increased standard deviation and strong mean bias. Surprisingly, the median bias is never strong in most of the cells.

The rescaled sampling error in transition probability parameter p_{21}

Next we consider the behaviour of the rescaled sampling error in p_{21}. In Figure 3.18 the main diagonal displays the distribution of ξ_{21} when states are not observed under the

Table 3.15: Summary statistics for the rescaled sampling error $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$ when states are not observed.

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	-0.0563	-0.1263	-0.448	-1.5654	-3.4586	-5.3302	-5.7882
Median	0.0147	0.0157	0.0284	-0.0713	-0.0801	-0.0853	-0.0572
Interquartile	1.3367	1.2483	1.1682	1.252	1.4145	1.3123	1.1576
Std	0.9906	0.9943	2.5839	6.6473	12.7759	19.4855	20.589
Skewness	-0.4763	-1.542	-11.9007	-6.73	-5.6728	-6.2991	-6.4197
Kurtosis	3.4574	11.8081	187.4717	55.298	41.6755	52.335	59.3035
Min	-4.1805	-9.6552	-44.9221	-73.4098	-146.0597	-220.4332	-283.0834
Max	2.6489	1.8858	1.4217	0.9537	0.7191	0.4506	0.2761
T=2,000							
Mean	-0.024	-0.0544	-0.1095	-0.4818	-2.2035	-4.75	-8.7758
Median	0.0577	0.0207	0.0469	0.044	-0.0544	-0.0792	-0.0762
Interquartile	1.2413	1.3633	1.2025	1.1263	1.1292	1.2881	1.4009
Std	0.964	0.9571	0.9548	3.5303	11.9097	19.7314	31.9542
Skewness	-0.5309	-0.4993	-1.0916	-15.8249	-8.0958	-6.464	-5.698
Kurtosis	3.608	3.0735	5.2549	325.1056	75.4821	54.5856	43.5069
Min	-4.4138	-3.7528	-5.7708	-82.8932	-140.1175	-261.7335	-356.5173
Max	2.5637	2.1154	1.8799	1.3156	0.9496	0.6552	0.4753
T=4,000							
Mean	-0.0514	-0.0754	-0.0883	-0.1918	-0.6913	-5.226	-9.7038
Median	-0.0137	0.0023	0.0274	0.0193	0.014	-0.1141	-0.1553
Interquartile	1.285	1.3146	1.3495	1.2991	1.1419	1.2985	1.7388
Std	1.0116	0.9857	0.9776	1.1395	6.7212	25.8314	39.3149
Skewness	-0.2976	-0.5253	-0.6899	-2.7095	-17.3268	-6.6447	-5.5202
Kurtosis	3.2164	3.7799	3.6891	20.1325	317.8163	51.1818	36.3957
Min	-3.4824	-5.0763	-4.3563	-11.732	-138.0518	-267.7412	-353.2354
Max	3.1062	2.3626	2.0943	1.6159	1.2327	1.2164	0.6151
T=8,000							
Mean	-0.062	-0.0911	-0.1331	-0.1606	-0.2317	-1.7301	-7.7418
Median	-0.0248	-0.0567	-0.0369	-0.0378	-0.0095	-0.0378	-0.1504
Interquartile	1.3299	1.2886	1.3082	1.2725	1.198	1.2026	1.3958
Std	1.0165	1.0156	1.0055	0.9581	1.1585	18.2802	49.489
Skewness	-0.3711	-0.2297	-0.4577	-0.7444	-3.7885	-21.8464	-9.5633
Kurtosis	3.4223	3.0989	3.3996	3.841	41.0975	534.5617	102.7791
Min	-4.4218	-3.4552	-3.8461	-4.7746	-16.0981	-484.64	-651.5851
Max	2.973	3.0176	2.6687	1.7944	1.5027	1.1514	0.8183
T=16,000							
Mean	-0.0594	-0.0699	-0.1055	-0.123	-0.1097	-0.3091	-2.4028
Median	-0.0271	-0.0589	-0.0615	-0.0331	-0.0052	-0.0056	-0.0721
Interquartile	1.2252	1.3006	1.3487	1.3037	1.2958	1.2258	1.2517
Std	0.951	0.9814	1.0052	0.9629	0.9452	1.5709	26.2608
Skewness	-0.0497	-0.2519	-0.2513	-0.3623	-0.741	-5.9556	-25.2797
Kurtosis	3.2193	3.2414	3.0212	2.9897	3.432	58.3969	713.1538
Min	-3.2464	-3.8713	-3.9226	-3.7004	-3.5339	-18.498	-763.3198
Max	2.9212	2.534	2.9538	2.1567	1.656	1.8234	1.1395
T=32,000							
Mean	-0.0248	-0.0581	-0.0882	-0.1387	-0.0974	-0.1498	-0.5771
Median	-0.0023	-0.0039	-0.0626	-0.0676	-0.0135	0.0202	0.0534
Interquartile	1.34	1.3213	1.3283	1.2948	1.2111	1.2284	1.1144
Std	0.9987	0.9697	0.9954	0.9625	0.9782	1.1906	8.2594
Skewness	-0.1941	-0.1415	-0.2062	-0.4031	-0.6769	-7.33	-27.2103
Kurtosis	3.3361	3.1195	3.2255	3.3584	3.8206	132.9369	790.7627
Min	-3.7145	-3.319	-4.1968	-4.2067	-4.8837	-22.8005	-246.0657
Max	3.2208	3.3776	3.0992	2.4646	2.2187	1.914	1.3661
T=64,000							
Mean	0.0175	-0.0315	-0.0617	-0.1095	-0.0536	-0.0584	-0.1064
Median	0.0671	0.0275	-0.0406	-0.0586	-0.0033	0.0518	0.0572
Interquartile	1.3282	1.2968	1.3963	1.3126	1.2544	1.2088	1.1882
Std	0.9958	1.004	0.9955	0.9917	0.9514	0.9287	0.9445
Skewness	-0.1421	-0.1561	-0.1329	-0.1153	-0.2934	-0.7446	-1.196
Kurtosis	3.1212	3.2065	2.7552	2.9862	3.2804	4.1234	6.4015
Min	-4.017	-3.7251	-3.117	-3.263	-3.5971	-5.0606	-5.8649
Max	2.8312	3.1619	2.7204	3.0436	2.6751	2.0159	1.9083

Figure 3.17: Distribution of the rescaled sampling error $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$ when states are not observed.

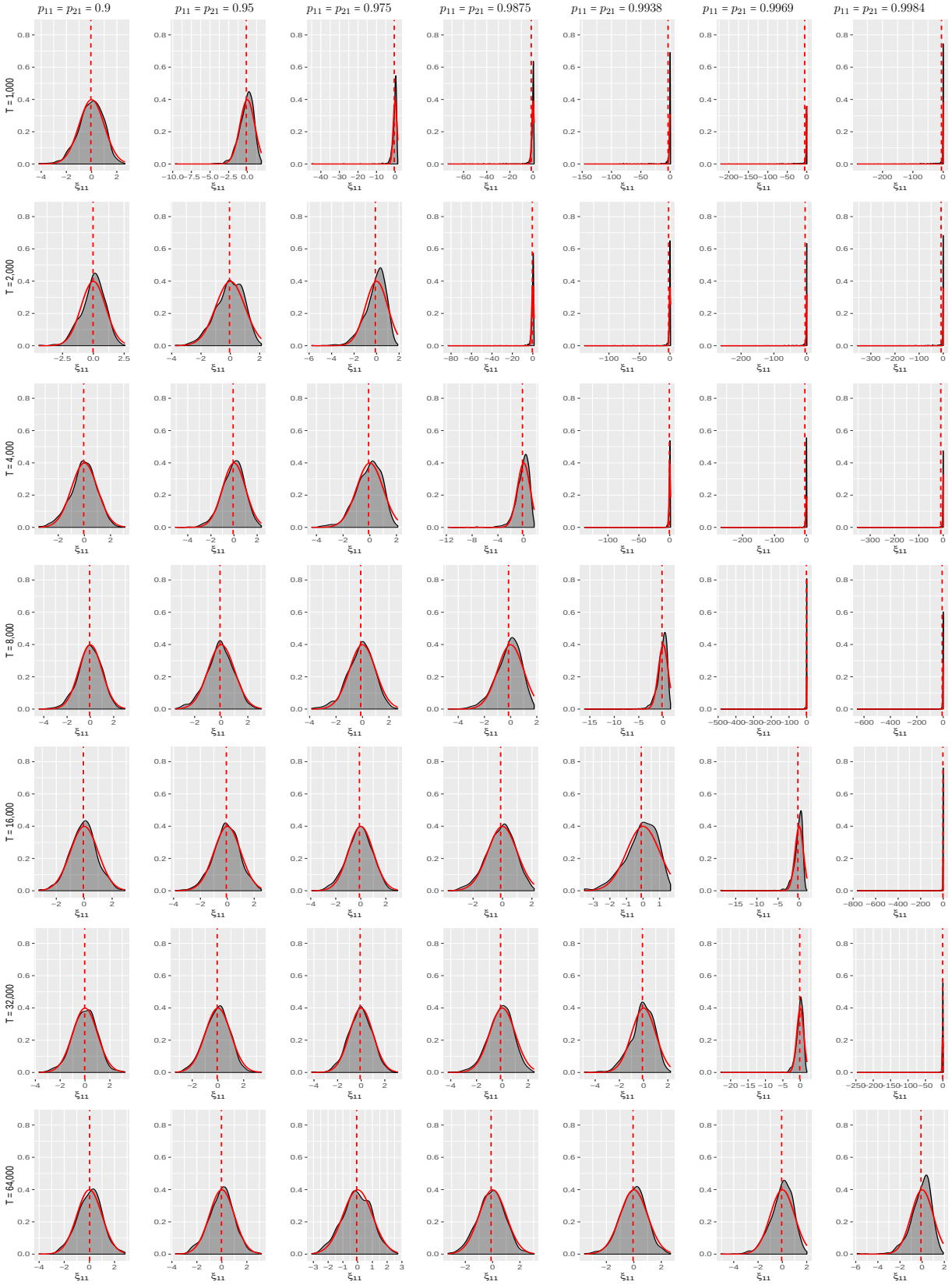


Table 3.16: Summary statistics for the rescaled sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are not observed.

	$p_{11} = p_{21} = 0.9$	$p_{11} = p_{21} = 0.95$	$p_{11} = p_{21} = 0.975$	$p_{11} = p_{21} = 0.9875$	$p_{11} = p_{21} = 0.9938$	$p_{11} = p_{21} = 0.9969$	$p_{11} = p_{21} = 0.9984$
T=1,000							
Mean	-0.1893	-0.3696	-0.5074	-0.439	-0.3139	-0.1784	-0.1104
Median	0.0314	0.1791	0.0652	0.019	0.0048	-0.0249	-0.0213
Interquartile	1.3383	0.9215	0.3168	0.136	0.1208	0.1055	0.0611
Std	0.9786	1.0681	1.6469	1.9073	1.5462	0.8861	0.7578
Skewness	-1.234	-2.5	-4.6594	-7.9734	-10.0978	-15.8916	-20.5374
Kurtosis	4.516	10.806	29.6972	83.2534	122.3703	332.1948	490.7491
Min	-4.4099	-7.2621	-14.7856	-29.0554	-24.4431	-21.0806	-19.7901
Max	1.0993	0.7609	0.2921	0.1409	0.0447	0.0121	0.0193
T=2,000							
Mean	-0.1316	-0.2992	-0.5049	-0.479	-0.362	-0.2286	-0.1507
Median	0.0715	0.1653	0.1009	0.0303	0.0089	0.002	-0.03
Interquartile	1.3698	0.9915	0.6106	0.2056	0.1265	0.1385	0.1022
Std	1.0402	1.0661	1.4715	1.7564	1.424	0.7495	0.5282
Skewness	-1.0317	-1.9189	-4.0218	-7.5687	-7.8856	-7.7216	-11.2435
Kurtosis	4.2783	6.8278	26.044	82.787	79.8384	84.4878	175.3221
Min	-5.0152	-6.3243	-14.1381	-27.0362	-18.5432	-11.367	-9.9867
Max	1.5137	0.8308	0.5056	0.2509	0.0643	0.0156	0.0198
T=4,000							
Mean	-0.0503	-0.1591	-0.4407	-0.5884	-0.4766	-0.4372	-0.2994
Median	0.0617	0.1658	0.157	0.0498	0.0147	0.004	0.0006
Interquartile	1.2631	1.1252	0.7472	0.2403	0.0474	0.1783	0.157
Std	0.9973	0.9697	1.3335	2.0711	2.027	2.1453	1.4132
Skewness	-0.7666	-1.6256	-3.3414	-7.6593	-9.7487	-12.8899	-11.7417
Kurtosis	3.6874	6.3211	17.1904	101.5993	139.036	211.3321	164.1825
Min	-4.1794	-6.3515	-10.0031	-36.4308	-37.4911	-42.6609	-22.5713
Max	1.9719	1.1969	0.5706	0.2152	0.1303	0.027	0.0098
T=8,000							
Mean	-0.0378	-0.1201	-0.3455	-0.4994	-0.5542	-0.4648	-0.4161
Median	0.0493	0.1038	0.2165	0.0794	0.0243	0.007	0.0019
Interquartile	1.2699	1.2685	0.9951	0.3607	0.0443	0.1022	0.1915
Std	0.9777	0.9736	1.1171	1.5149	2.1297	2.3456	2.727
Skewness	-0.5229	-1.0799	-2.3201	-3.7431	-7.2527	-14.9384	-22.5082
Kurtosis	3.3347	3.9227	9.6468	19.9597	73.2803	301.105	599.4694
Min	-3.6171	-3.9508	-7.6755	-13.5907	-29.7585	-54.2479	-75.9934
Max	2.2424	1.452	0.6489	0.3489	0.1208	0.0473	0.0219
T=16,000							
Mean	-0.0195	-0.0495	-0.2217	-0.4264	-0.5837	-0.4678	-0.4338
Median	0.0391	0.0841	0.2091	0.133	0.0401	0.0114	0.0032
Interquartile	1.3137	1.3249	1.0845	0.6001	0.0946	0.0723	0.0481
Std	0.9883	0.9735	1.0797	1.3754	2.4197	1.7741	2.098
Skewness	-0.4749	-0.8092	-2.0114	-4.028	-7.5376	-7.1616	-13.4301
Kurtosis	3.2032	3.6773	7.8031	27.469	73.6	64.4645	226.3301
Min	-3.5674	-3.9381	-6.4101	-15.652	-31.0252	-20.2575	-42.0843
Max	2.3721	1.8629	1.1831	0.4026	0.1439	0.0593	0.0244
T=32,000							
Mean	-0.0356	-0.0267	-0.1443	-0.3792	-0.5433	-0.4789	-0.4982
Median	0.0447	0.0941	0.1898	0.1903	0.0641	0.0184	0.0052
Interquartile	1.3696	1.3231	1.2316	0.809	0.2992	0.0531	0.0039
Std	1.0245	1.0476	1.0859	1.2455	1.813	1.6962	2.4104
Skewness	-0.3558	-0.8132	-1.7328	-3.2885	-5.3718	-6.0945	-11.0289
Kurtosis	3.049	4.0516	6.7915	18.745	41.4748	50.5622	156.9684
Min	-3.4542	-4.8292	-6.0863	-10.8867	-20.3003	-20.4111	-43.3808
Max	2.8021	2.083	1.2598	0.6217	0.3004	0.171	0.0468
T=64,000							
Mean	-0.0391	0.0015	-0.0606	-0.2674	-0.5268	-0.5973	-0.476
Median	0.0084	0.0947	0.1408	0.1688	0.1019	0.0322	0.009
Interquartile	1.3694	1.4237	1.2894	1.0274	0.5944	0.0259	0.0059
Std	1.0278	1.0127	0.9937	1.0748	1.4684	2.2672	2.3701
Skewness	-0.2561	-0.3447	-1.121	-2.9005	-3.093	-6.4552	-9.3714
Kurtosis	2.9975	2.745	4.4756	18.9468	14.1069	54.9377	106.1803
Min	-3.7943	-3.2563	-5.0878	-11.4335	-11.9519	-24.9104	-32.8855
Max	2.648	2.2684	1.4188	0.7008	0.3393	0.1256	0.0369

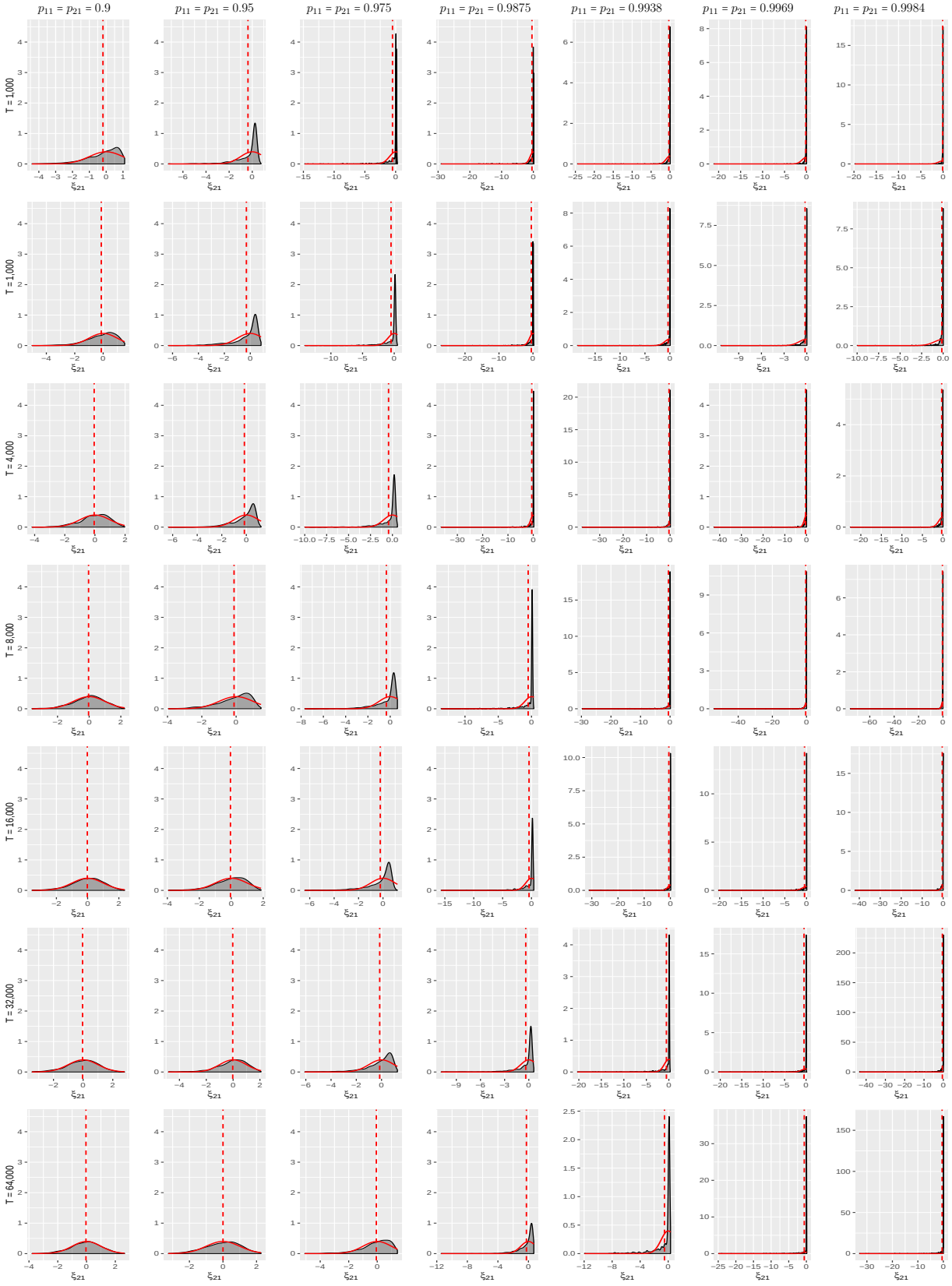


Figure 3.18: Distribution of the rescaled sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are not observed.

alternative asymptotic sequence Case 1. It can be observed that, as the run length T doubles, the behaviour deteriorates very rapidly. From Table 3.16 it can be noted that both skewness and kurtosis increase quickly, accompanied by a strong mean bias and increasingly large negative values. The standard deviation also increases but less rapidly, and the median bias remains low for most of the cells on the main diagonal. Again, establishing the limiting distribution would involve non-trivial derivations. However, it is reasonable to assume that the limiting distribution will not be the standard normal distribution.

The usual asymptotics in the unobserved case work well, but as it is evident from the last column in Figure 3.18, a much larger sample size would be required for the behaviour to get close to the reference distribution for that particular specification of the P matrix.

Next, we can examine the behaviour of the rescaled sampling error by moving across the columns. It can be seen that the matrix G does not need to get very close to zero for the behaviour of the rescaled sampling error to start to deteriorate. Examination of the first plot in Figure 3.18 for the run length $T = 1000$, reveals pronounced distortions as compared to the observed case in Figure 3.14. As we move to the next plot across the columns, the behaviour gets substantially distorted away from the reference distribution, whilst in the observed case we could observe a much milder relative distortion. Summary statistics in Table 3.16 indicate a strong negative skewness, high kurtosis and a sizeable mean bias equal to *0.37*.

In summing up the simulation results, the lack of observability of the states seems to add an additional layer of distortions which affect the behaviour of rescaled sampling errors in the parameters of interest. First, the simulation results reveal that the parameter estimates in the high occupancy *state 1* are negatively affected as a result of not being able to observe the states. This implies that the behaviour in high occupancy state is not insensitive to what happens in the low occupancy state. That is, the stronger distortions are in the low occupancy *state 2*, the more striking the "contamination" effect is in the high occupancy *state 1*. Second, the inability to perfectly distinguish one state from another means that samples of considerable size are required for the behaviour of the standardised sampling

errors to be well approximated by the usual asymptotic theory. In particular, skewness present in the rescaled sampling error in θ_1 under the alternative asymptotic sequence Case 1 washes away with run lengths larger than $T = 1000$. For the rescaled sampling error in θ_2 the notable distortions do not seem to wash away as the run length increases, because the expected number of time periods spent in *state 2* remains stochastically bounded. However, if we move sufficiently down to a lower sub-diagonals, meaning that for each specification of P we have a larger sample, then the behaviour under the alternative asymptotic sequence Case 1 is reasonably well approximated by the standard asymptotic theory. There are also distortions present in the rescaled sampling error in p_{11}, but it remains unclear whether they will wash away further down the sequence. Having said that, the plots seem to suggest increasing negative skewness and kurtosis. Again, if we move to a lower sub-diagonals then the behaviour improves, although it is still unclear what will be the limiting distribution. For the rescaled sampling error in p_{21} we do not expect distortions to go away regardless of the sample size, because the expected number of state transitions from *state 2* to *state 1* remains stochastically bounded, and at the same time P is getting closer to a boundary. Although we cannot tell with certainty the limiting distribution it will converge to, it is quite conceivable that it will not be the standard normal distribution.

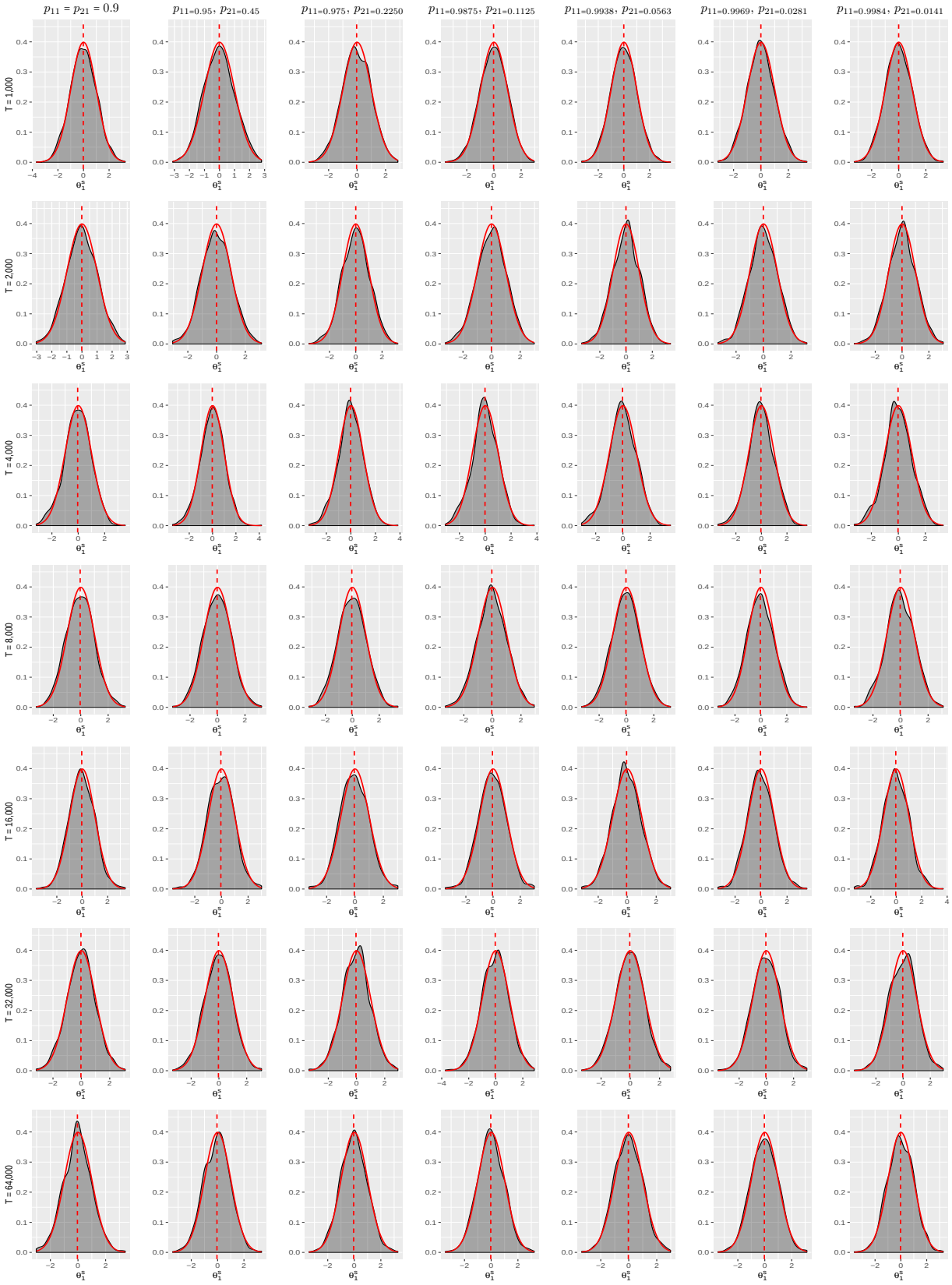
3.3.2 Alternative asymptotic sequence Case 2

In this section we report results for the alternative asymptotic sequence Case 2 described by specification in (3.46). Under this asymptotic sequence, the parameter values of the transition probability matrix change with the run length in a fashion that results in the expected number of time periods spent in both regimes proportional to the run length. This implies that as the run length increases one spends on average longer time periods in both states. The consequence of this characteristic on the behaviour of the rescaled sampling errors is interesting, and it will be analysed next for both observed and unobserved scenarios.

Table 3.17: Summary statistics for the rescaled sampling error $\theta_1^s = (\hat{\theta}_1 - \theta_1)/se(\hat{\theta}_1)$ when states are observed.

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	-0.0259	-0.0157	-0.004	-0.0225	-0.0233	-0.017	-0.0167
Median	-0.0441	-0.0306	-0.0409	-0.0076	-0.0056	-0.0158	-0.0317
Interquartile	1.3783	1.3633	1.412	1.422	1.4007	1.442	1.4148
Std	1.0217	1.0134	1.0112	1.0125	1.0017	1.0174	1.0048
Skewness	0.0208	-0.0418	-0.0506	-0.0263	-0.0196	-0.0208	0.0349
Kurtosis	2.8379	2.7422	2.6846	2.6474	2.6409	2.7103	2.8402
Min	-3.0392	-3.1212	-2.7517	-3.2135	-2.7031	-3.0253	-3.0092
Max	2.9356	2.746	2.8168	2.9592	2.7176	3.3447	3.2658
T=2,000							
Mean	-0.0128	-0.0022	0.0000	-0.011	-0.0074	-0.0143	-0.0057
Median	-0.0249	-0.0292	0.0128	0.036	0.0341	-0.0095	0.0243
Interquartile	1.4135	1.417	1.4008	1.3735	1.3897	1.3733	1.4179
Std	1.0224	1.0345	1.0374	1.0231	1.0228	1.0256	1.0321
Skewness	-0.0031	-0.0095	-0.0848	-0.1187	-0.0933	-0.0598	-0.0303
Kurtosis	2.8333	2.8982	2.9464	2.8669	2.9118	2.9405	2.9917
Min	-3.0294	-3.0587	-3.5622	-3.2344	-3.3857	-3.3189	-3.5629
Max	2.8661	3.127	3.2029	2.979	3.3883	3.1587	3.0667
T=4,000							
Mean	-0.0436	-0.0255	-0.0236	-0.021	-0.0271	-0.0397	-0.0218
Median	-0.0285	0.0157	-0.0316	-0.0185	-0.0329	-0.0606	-0.0333
Interquartile	1.3196	1.3434	1.3035	1.277	1.3096	1.2846	1.2732
Std	1.0076	1.0265	1.0186	1.0025	1.0011	1.0133	1.0105
Skewness	-0.1627	-0.1309	-0.1146	-0.1048	-0.1097	-0.0848	-0.0853
Kurtosis	3.0525	3.2294	3.1669	3.079	3.0282	3.1141	3.0424
Min	-3.1814	-3.4241	-3.4161	-3.0973	-3.0353	-3.4053	-3.2639
Max	3.5566	4.1956	3.8281	3.7891	3.501	3.5573	3.323
T=8,000							
Mean	-0.0414	-0.0273	-0.0379	-0.042	-0.0356	-0.0437	-0.0257
Median	-0.058	-0.023	-0.019	-0.0505	-0.0315	-0.0532	-0.0299
Interquartile	1.3972	1.371	1.4064	1.3927	1.3693	1.4316	1.4133
Std	1.0241	1.0171	1.0344	1.0247	1.0203	1.0435	1.028
Skewness	0.0825	0.0396	0.0141	-0.0335	0.0062	0.0283	-0.0196
Kurtosis	2.9428	2.9261	2.8492	2.937	2.8947	2.882	2.8869
Min	-3.2448	-3.268	-3.2204	-3.2988	-3.2818	-3.3153	-3.3541
Max	3.2442	3.2043	3.3816	2.9237	3.2264	3.5496	3.0955
T=16,000							
Mean	-0.0392	-0.0436	-0.049	-0.0448	-0.0413	-0.0419	-0.0366
Median	-0.0457	-0.0252	-0.0539	-0.0599	-0.0958	-0.081	-0.0699
Interquartile	1.3374	1.3971	1.3878	1.3438	1.3149	1.3695	1.3916
Std	1.0028	1.0071	1.0023	1.0028	0.9842	1.0112	1.0046
Skewness	0.0715	0.0056	0.0647	0.0001	0.0335	0.0921	0.0187
Kurtosis	3.208	3.1581	3.1103	3.1607	3.0359	3.1882	3.121
Min	-3.6168	-3.6645	-3.2458	-3.514	-3.1446	-3.3168	-3.2946
Max	3.3799	3.0068	3.0235	3.0241	2.9905	3.5243	3.6894
T=32,000							
Mean	-0.0285	-0.0346	-0.0375	-0.038	-0.0364	-0.0369	-0.0268
Median	0.0074	-0.0368	-0.0101	0.0047	-0.0362	-0.026	0.0086
Interquartile	1.3027	1.3754	1.2632	1.3783	1.3371	1.3644	1.3955
Std	1.0004	1.0012	0.9899	0.9969	0.9787	0.996	0.9962
Skewness	-0.0359	-0.0458	-0.047	-0.07	-0.0363	-0.0505	-0.0512
Kurtosis	3.0672	3.0172	3.1061	3.0992	3.0095	3.0714	2.9846
Min	-3.1782	-3.3337	-3.3716	-3.7367	-3.33	-3.5555	-3.5919
Max	3.0913	3.067	2.9292	2.8461	2.7531	2.9773	2.9479
T=64,000							
Mean	-0.0398	-0.0493	-0.0458	-0.0315	-0.0277	-0.0306	-0.0231
Median	-0.0377	-0.0128	-0.043	-0.0437	-0.0282	-0.0375	-0.0309
Interquartile	1.3578	1.4007	1.3575	1.338	1.4151	1.3967	1.3917
Std	1.0173	1.0035	0.9988	0.9918	0.9979	1.008	1.0043
Skewness	0.007	-0.0267	0.0344	0.0285	0.0364	-0.0149	-0.0241
Kurtosis	3.1573	3.0231	3.0309	3.1084	2.9746	2.9626	3.0002
Min	-3.0212	-3.4194	-3.4131	-3.4362	-3.3675	-3.4759	-3.6106
Max	3.4186	3.3348	3.2812	3.2191	2.9852	3.1265	3.2827

Figure 3.19: Distribution of the rescaled sampling error of $\theta_1^s = (\hat{\theta}_1 - \theta_1)/se(\hat{\theta}_1)$ when states are observed.



Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000.

Observed states

First we consider a world in which we could observe the states. As we have seen earlier, the consequence of this assumption is that the estimation procedure is much more straightforward, the behaviour of the rescaled sampling errors is relatively better, and it is considerably easier to pin down the limiting behaviour. When states can be observed, we know with absolute certainty in which state we are in, and hence, the copula parameter estimates do not depend on the transition parameters p_{11} and p_{21}. In what follows, we present the empirical results from Monte Carlo simulations when states can be fully observed.

The rescaled sampling error in copula parameter θ_1 in state 1

Figure 3.19 displays the sampling distribution of θ_1^s in the high occupancy state. Along the main diagonal we can observe the behaviour of the rescaled sampling error under the alternative asymptotic sequence Case 2 described in (3.46). That is, the G matrix is held constant whilst the run length T and the probability transition matrix P are allowed to vary. As one moves down the main diagonal, the heuristic theory would suggest that the distribution of the rescaled sampling error would converge to the standard normal distribution. The visual inspection of the plots and the summary statistic in Table 3.17 seem to suggest that this is indeed the case. The copula parameters from *state 1* are being well determined, and the usual asymptotics seem to work reasonably well in all plots on the main diagonal. This is not surprising as one moves down the main diagonal by one cell, the run length T doubles, and the number of time periods spent in *state 1* also doubles. Therefore, a large enough sample from *state 1* can be obtained by moving sufficiently far down the asymptotic sequence. In addition, the distribution of the sampling error in θ_1 seems to be shrinking at the rate $T^{-\frac{1}{2}}$, which is broadly in line with what the heuristic theory would suggest.

We can also analyse the behaviour under the standard asymptotic sequence by moving vertically down the column. Under this regular asymptotic sequence, the matrix P is held fixed whilst the run length T and matrix G vary. This results in the expected number

of time periods spent in *state 1* to double as the run length T doubles. Therefore, the information about the copula parameter θ_1 in *state 1* would increase to infinity, and the distribution of the rescaled sampling error would converge to the standard normal distribution. The plots in Figure 3.19 and the summary statistics in Table 3.17 indicate that the behaviour of the rescaled sampling error is reasonably well approximated by the reference standard normal distribution.

We could also analyse how the behaviour of the rescaled sampling error changes as G gets close to zero for a given run length T. This is the behaviour observed as one moves along any row across columns. This would also imply that for a given fixed T the transition probability matrix P would vary, resulting in the expected number of time periods spent in *state 1* to remain constant. In particular, in the first plot for $T = 1000$, the expected number of time periods spent in *state 1* is 900, which remains the same as we move across columns. However, the expected number of transitions from *state 1* to *state 1* increases by $\frac{q_{22}}{10}$, using equation (3.51). In other words, the steady state probability of being in *state 1* remains constant, and once the process is in *state 1*, it will tend to remain in that state longer. Furthermore, as a result of an increasing standard deviation of the number of transitions from *state 1* to *state 1*, there will be some sample realisations providing more information about the copula parameter θ_1, and some providing less information, but on average it will be sufficiently high, that is 900. Therefore, it is expected that the behaviour of the rescaled sampling error in θ_1 will be reasonably well approximated by the standard normal distribution. The plots in Figure 3.19 indicate that there are no noticeable discrepancies between the rescaled sampling error in θ_1 and the superimposed reference distribution. The summary statistics in Table 3.17 indicate that the standard error, interquartile range, skewness and kurtosis remain close to the reference distribution. In addition, summary statistics suggest that the MLE of θ_1 is mean and median unbiased.

The rescaled sampling error in copula parameter θ_2 in state 2

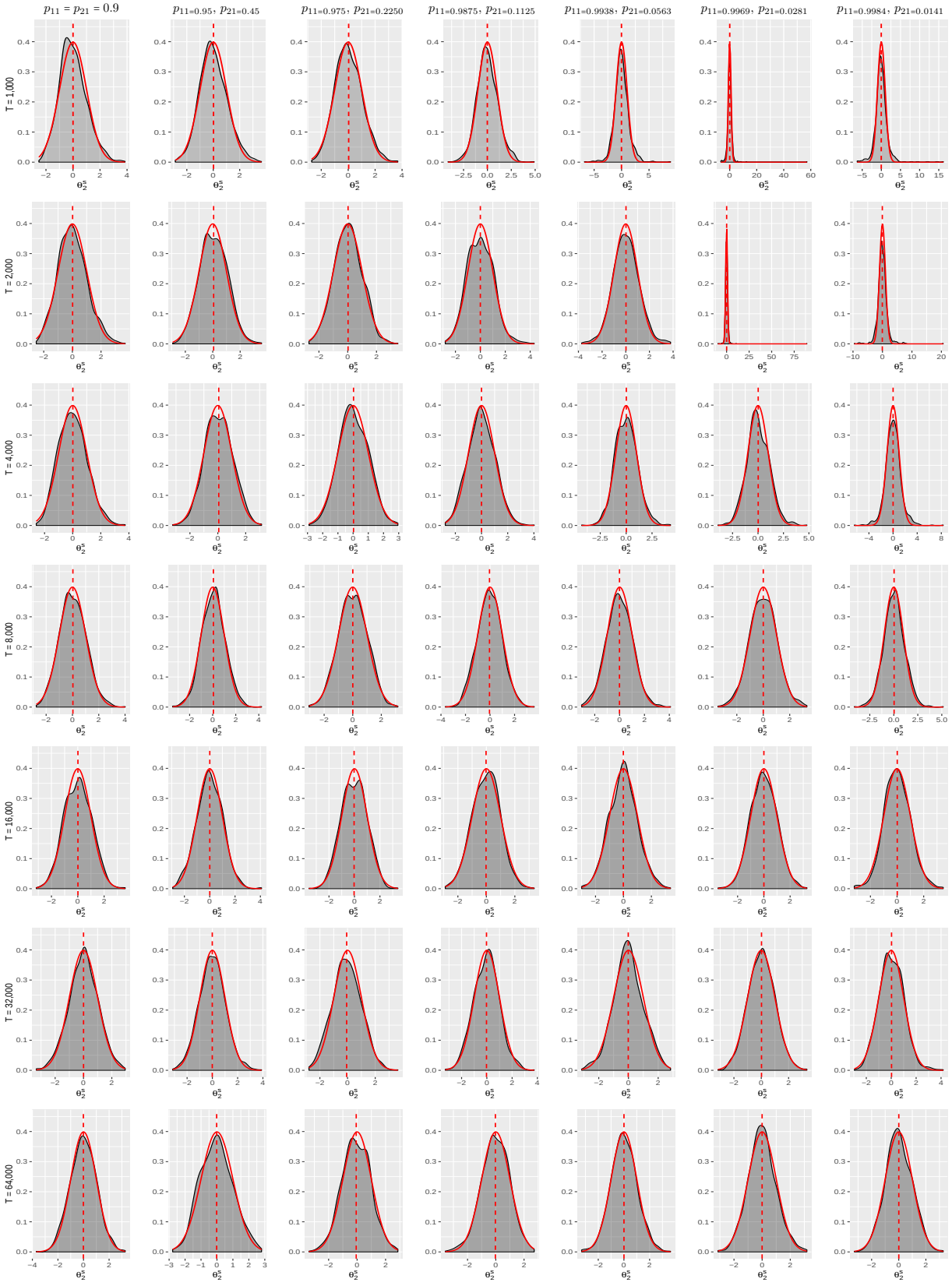
Next we consider the distribution of the rescaled sampling error in θ_2 associated with *state 2*. Figure 3.20 displays the sampling distribution of θ_2^s in *state 2*. The behaviour along the main diagonal seems to be consistent with the heuristic theory, which suggests

Table 3.18: Summary statistics for the rescaled sampling error $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$ when states are observed.[1]

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	0.0022	0.0195	0.0346	-0.0102	-0.0173	0.0276	-0.0234
Median	-0.0805	-0.0741	-0.0384	-0.0559	-0.068	-0.0872	-0.1074
Interquartile	1.289	1.3891	1.3839	1.4231	1.4529	1.4775	1.5386
Std	0.9928	1.0293	1.0057	1.0999	1.3241	2.5322	1.7149
Skewness	0.4497	0.2987	0.2914	0.1754	0.2895	16.0983	2.3071
Kurtosis	3.2786	3.1222	3.0732	3.5472	7.181	399.1133	23.417
Min	-2.5225	-2.8721	-2.7108	-4.1221	-6.7582	-6.5598	-6.2941
Max	3.847	3.6409	3.6816	4.8946	8.9171	62.9695	16.1848
T=2,000							
Mean	0.0165	0.0645	0.0236	-0.0007	0.0182	0.079	-0.0449
Median	-0.0288	0.0295	0.022	-0.0193	-0.0106	-0.0446	-0.1127
Interquartile	1.3623	1.387	1.3043	1.5006	1.4156	1.5273	1.578
Std	1.0085	1.007	0.9709	1.0565	1.1023	3.1179	1.7075
Skewness	0.3277	0.1212	0.08	0.2748	0.1138	23.303	1.7134
Kurtosis	3.0401	2.8653	2.9277	3.2613	3.4884	666.1515	29.0591
Min	-2.5208	-2.9266	-2.6946	-2.8135	-3.7304	-9.5582	-9.5582
Max	3.6891	3.5598	3.4862	4.3225	3.7842	89.0523	20.6222
T=4,000							
Mean	0.0125	0.0653	0.0275	-0.0015	0.0187	0.0182	0.0031
Median	-0.0284	0.0309	-0.0174	-0.0444	0.0136	-0.0509	-0.0083
Interquartile	1.4018	1.3925	1.3624	1.3981	1.4625	1.4546	1.4893
Std	1.0037	1.0005	0.9653	1.0294	1.0822	1.1163	1.2973
Skewness	0.299	0.033	0.088	0.1222	0.1888	0.2213	0.2371
Kurtosis	2.9486	2.8324	2.8026	2.9683	3.4615	3.5413	6.4771
Min	-2.5889	-3.243	-2.9053	-2.7674	-4.2996	-3.9324	-6.4284
Max	3.7365	3.1367	2.9454	4.024	4.2879	4.7994	8.3161
T=8,000							
Mean	0.0338	0.0589	0.0054	-0.0417	-0.0093	0.0014	0.0311
Median	-0.0158	0.0772	0.0096	-0.048	-0.0388	0.0055	0.0407
Interquartile	1.3413	1.3706	1.3818	1.3338	1.377	1.3985	1.3847
Std	1.0084	1.0268	0.9923	1.0015	1.0527	1.0444	1.0904
Skewness	0.1865	0.007	-0.0222	0.0297	0.0828	0.0414	0.0818
Kurtosis	2.9969	3.2375	2.8123	2.9715	3.1572	3.1324	3.6981
Min	-2.7498	-3.4764	-3.0638	-3.6479	-3.0551	-3.4249	-4.1133
Max	4.0113	4.2386	3.1597	3.6019	4.0984	3.2969	5.1488
T=16,000							
Mean	0.01	0.0042	-0.0279	-0.0225	-0.008	0.0016	0.0142
Median	0.021	-0.0067	-0.044	0.0197	0.0067	-0.0051	-0.0067
Interquartile	1.4854	1.3849	1.3983	1.3451	1.3403	1.4057	1.3401
Std	1.0372	1.0219	1.0016	0.9702	1.0068	1.0003	0.9946
Skewness	0.0616	0.0717	-0.0093	-0.0373	0.0773	0.0347	0.0881
Kurtosis	2.8417	3.0979	2.7256	2.9359	3.0503	3.0093	3.4265
Min	-3.1526	-2.9301	-3.573	-2.8844	-2.9947	-3.4958	-3.2106
Max	3.6007	4.092	3.4092	3.3597	3.3711	3.2766	3.4652
T=32,000							
Mean	0.0041	-0.0064	-0.0616	-0.0268	-0.0146	0.0181	0.0264
Median	0.0111	-0.0048	-0.064	-0.0003	-0.0347	0.0326	-0.0087
Interquartile	1.3653	1.361	1.39	1.3884	1.282	1.3412	1.3473
Std	1.0259	1.0397	1.0375	1.0084	0.9949	0.9754	0.9895
Skewness	-0.077	0.1052	0.0784	0.0562	-0.0106	0.0747	0.2435
Kurtosis	3.0131	3.1745	2.8054	3.0731	3.04	2.9917	3.3679
Min	-3.3324	-3.1464	-2.7893	-3.2913	-3.1533	-3.1689	-2.9629
Max	2.9448	3.8927	3.6356	3.749	2.8317	3.3389	4.1745
T=64,000							
Mean	-0.0099	-0.0188	-0.0607	-0.0281	0.0033	-0.0026	0.0043
Median	0.0052	-0.0223	-0.0893	-0.0214	-0.0138	-0.0241	-0.0182
Interquartile	1.3781	1.4274	1.3507	1.3585	1.3486	1.2344	1.3488
Std	1.0171	1.0123	1.0031	0.998	1.0192	0.9859	0.9796
Skewness	0.0141	0.1522	-0.0038	-0.1914	0.0441	0.1617	0.0684
Kurtosis	2.9655	2.6785	2.954	3.1169	3.2344	3.1907	2.9449
Min	-3.6939	-2.7812	-3.3811	-3.576	-3.2956	-3.1396	-3.278
Max	3.2602	2.7958	2.8631	2.7204	3.632	3.1799	3.2792

Note: [1] Summary statistics are conditional on spending a positive number of time periods in *state 2*.

Figure 3.20: Distribution of the rescaled sampling error $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$ when states are observed.[1]



Note: [1] The distribution is conditional on spending a positive number of time periods in State 2. The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000.

that the distribution of the rescaled sampling error would converge to the standard normal distribution. This is because under the alternative asymptotic sequence Case 2, the number of time periods spent in *state 2* also doubles when the run length T doubles and, therefore, the expected number of time periods spent in *state 2* increases to infinity as the run length T goes to infinity. Hence, the distribution of an appropriately rescaled sampling error *state 2* is going to settle down. The rate $T^{-\frac{1}{2}}$ seems to be broadly in line with what the heuristic theory would suggest. The copula parameter θ_2 from *state 2* is being well determined, and the usual asymptotic theory works reasonably well in all plots on the main diagonal.

The behaviour under the standard asymptotics is described by each column, where both the run length T and matrix G increase whilst matrix P is being held constant. This implies that the total number of time periods spent in *state 2* is going to infinity, and hence the proportion of time periods spent in *state 2* is going to a constant. As T increases to infinity, the behaviour of the rescaled sampling error is starting to be better and better approximated by its limiting standard normal distribution. From Table 3.18 we can notice that a sizeable positive skewness present in the first cell for $T = 1000$ is gradually washed away with larger T.

We can also analyse how things behave as we let matrix G get very close to zero, whilst keeping the run length T constant. This is achieved by moving across the columns along each row. If one fixes T and varies G, this will also induce a change in P in such a way that the expected number of state transitions will halve with each move across the columns. This is because the matrix P is getting closer and closer to the boundary case presented in (3.46). However, the expected number of time periods spent in *state 2* is remaining constant. In particular, in the first plot for $T = 1000$, the expected number of time periods spent in *state 2* is 100, and the expected number of state transitions is 180, using (3.49) and (3.50) respectively. In the next cell, the expected number of time periods in *state 2* remains at 100, but the expected number of state transitions from *state 2* to *state 1* drops to 90. The steady state probability of being in *state 2* remains constant, but the expected number of transitions from *state 2* to *state 2* increases by $\frac{g_{22}}{10}$, using equation (3.52). However, as it was mentioned previously, the expected number of transitions from

state 1 to *state 1* also increases by the same amount $\frac{g_{22}}{10}$. Therefore, we would expect the standard deviation of the number of time periods spent in each state to increase as G gets close to zero. This will result in some sample realisations providing more information about the copula parameter θ_2, some providing very little information, and some providing no information at all. The explanation of why there will be some sample realisations with zero time periods spent in *state 2*, is because the expected number of time periods of 100 spent in state *state 2* remains the same, and it is considerably lower than the expected number of time periods spent in *state 1*, which is 900. Hence, given a small sample size T = 1000 and increased standard deviation of the number of time periods spent in each state, it is becoming more and more likely to obtain a sample realisation with zero time periods spent in *state 2* as matrix G gets closer to zero.

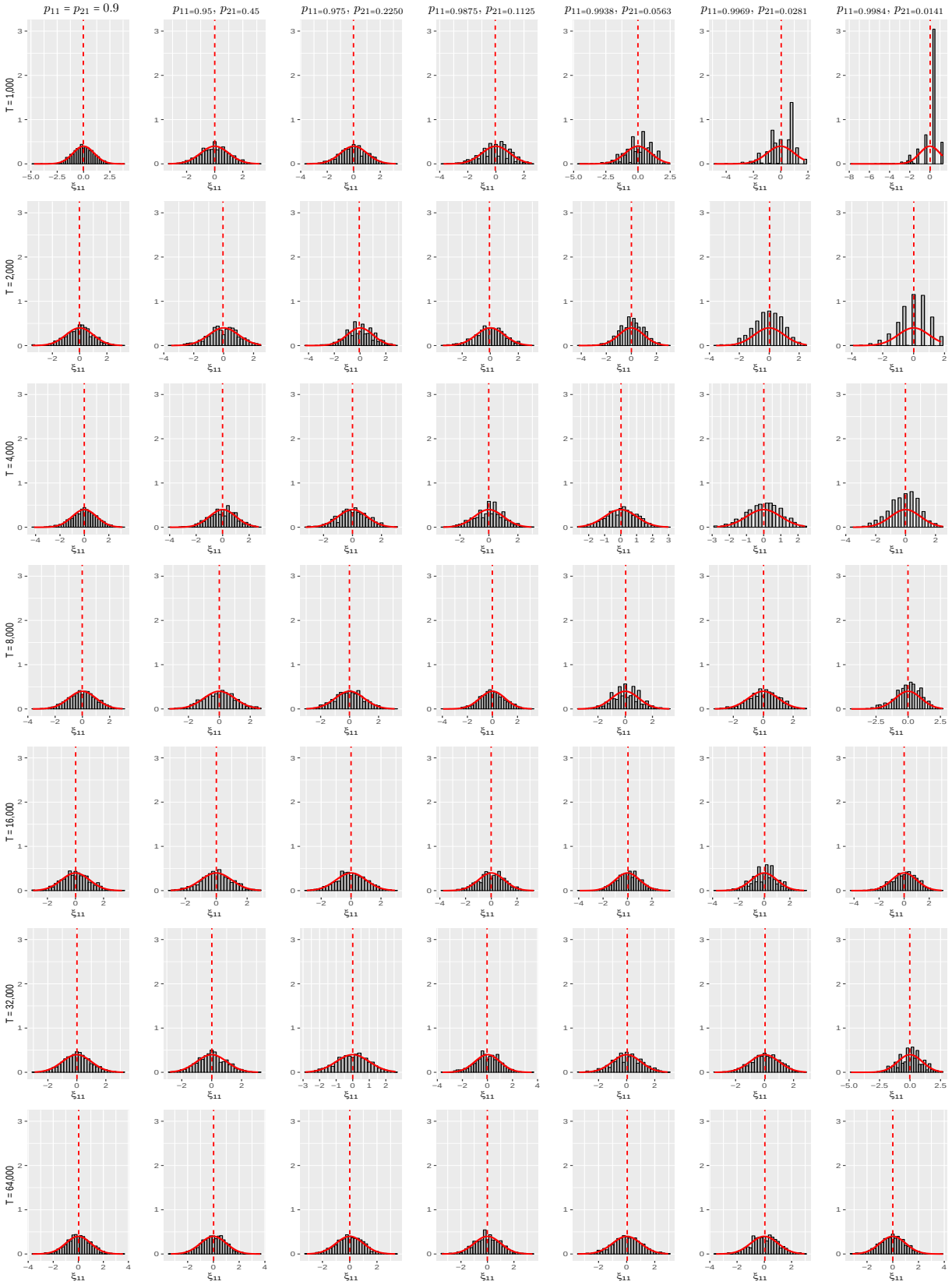
The plots in the upper right region in Figure 3.20 reflect these features in the behaviour of the rescaled sampling error. It can be noticed that the centre of the distribution is reasonably well approximated by the standard normal distribution, whilst there are also observations that lie considerably far down in the tails, where the reference distribution would assign a very tiny probability. These are the observations that have been obtained from those realisations with a small number of time periods spent in *state 2*, and hence they contain more noise and uncertainty. The mixture of reasonably decent estimates and poorly determined estimates seems to result in the usual asymptotics not working very well. The summary statistics in Table 3.18 indicate that the degree of skewness and kurtosis of the standardized sampling error are associated with the value of G. The closer the matrix G is to zero, the more prominent distortions in the behaviour of the standardized sampling error become. Moreover, the smaller is the sample size T, the more rapidly the behaviour deteriorates. The plots in the first two rows in Figure 3.20 and the corresponding cells in Table 3.18 reveal distinguishable strong distortions in the behaviour of the rescaled sampling error starting from sixth cell onwards. In addition, the median becomes biased, whilst the mean is never strongly biased in most of the cells. The distortions are in the form of noticeable positive skewness and strong kurtosis [9].

[9] Under the alternative asymptotic sequence 2 there are also replications in which there are no observations from *state 2*. We know that whatever estimate of θ_2 is obtained it will be inherently spurious. For

Table 3.19: Summary statistics for the rescaled sampling error $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$ when states are observed.

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	-0.0489	-0.0648	-0.0761	-0.0682	-0.0402	-0.0363	-0.0653
Median	-0.0558	-0.0349	-0.0295	0.0000	0.1412	-0.0312	0.3121
Interquartile	1.3226	1.4481	1.4084	1.393	1.4257	1.2632	0.974
Std	1.0133	1.0409	1.0196	0.9953	1.0242	1.0306	1.0237
Skewness	-0.1892	-0.2177	-0.2213	-0.3037	-0.704	-0.7519	-1.0669
Kurtosis	3.4539	2.8376	3.0417	2.9126	3.8373	3.6076	4.8327
Min	-4.7001	-3.5784	-3.7089	-3.0963	-4.392	-4.349	-5.8268
Max	3.3173	2.9314	3.2175	2.2652	2.378	1.6788	1.1862
T=2,000							
Mean	-0.0196	-0.0463	-0.0774	-0.0634	-0.0134	-0.01	-0.0357
Median	0.0079	-0.0081	-0.0602	-0.016	0.0387	-0.0056	-0.033
Interquartile	1.2864	1.3589	1.3805	1.3016	1.3453	1.3714	1.205
Std	0.9974	1.0201	1.0322	0.9784	0.9934	0.994	0.9711
Skewness	-0.1638	-0.2087	-0.1587	-0.2474	-0.3352	-0.4268	-0.6938
Kurtosis	3.2607	2.962	2.9572	3.0944	3.3225	3.1906	3.6535
Min	-3.4643	-3.3337	-3.7768	-3.3411	-3.4923	-4.1677	-4.3843
Max	3.1325	2.6558	2.9252	2.8043	2.8234	2.3748	1.6779
T=4,000							
Mean	-0.0103	-0.0127	-0.0185	-0.0305	0.009	-0.0218	-0.0425
Median	0.0111	0.0781	0.0053	0.0419	0.0632	0.0526	0.0061
Interquartile	1.3562	1.4477	1.4583	1.3386	1.381	1.4151	1.4116
Std	1.0395	1.0435	1.0386	1.0092	1.0059	1.0155	1.0173
Skewness	-0.1912	-0.2027	-0.1381	-0.185	-0.1903	-0.351	-0.4912
Kurtosis	3.1878	2.9321	2.7675	2.9236	2.8833	2.9705	3.2591
Min	-4.1198	-3.7901	-2.9738	-3.4066	-3.0533	-3.7796	-3.7472
Max	3.1183	2.6772	2.9007	2.8136	2.9869	2.8126	2.3733
T=8,000							
Mean	-0.0122	-0.016	-0.0267	-0.0229	0.0222	-0.0125	-0.0397
Median	0.0138	0.0081	-0.0132	-0.0133	0.0664	0.0285	0.0431
Interquartile	1.3897	1.3868	1.4107	1.3315	1.4049	1.3631	1.3639
Std	1.0208	1.04	1.0074	1.0139	1.0117	1.0327	1.019
Skewness	-0.0649	-0.096	-0.0158	-0.0663	-0.1604	-0.3148	-0.5049
Kurtosis	2.8404	2.945	2.7497	2.9508	3.015	3.1533	3.5505
Min	-3.6513	-3.7604	-3.0563	-3.5888	-3.4235	-3.5552	-4.7998
Max	2.9456	2.7899	3.2222	2.9805	3.0203	2.8838	2.5334
T=16,000							
Mean	0.0131	-0.0034	0.0006	-0.0206	0.0271	0.0109	-0.0155
Median	0.0097	0.0019	0.0074	0.0347	0.0383	0.0535	0.0589
Interquartile	1.3066	1.3303	1.3637	1.4261	1.4094	1.2666	1.3583
Std	0.9741	1.0029	0.9985	1.032	1.0196	1.002	1.0287
Skewness	0.0327	-0.0942	-0.0962	-0.179	-0.0801	-0.2555	-0.4483
Kurtosis	2.8894	2.9977	2.7989	2.9813	3.2856	3.182	3.5493
Min	-3.0158	-3.1363	-2.9608	-3.6954	-4.1504	-3.6145	-5.1877
Max	3.3105	2.9238	2.6752	3.2521	3.3776	3.0825	2.7687
T=32,000							
Mean	0.0089	0.0041	0.0066	-0.0311	0.0387	0.0105	-0.0315
Median	-0.0108	-0.0169	0.008	0.0119	0.0441	0.0562	-0.0112
Interquartile	1.2435	1.3039	1.3604	1.377	1.3837	1.4045	1.3088
Std	0.9537	0.9827	0.9758	1.0464	1.066	1.0232	1.0195
Skewness	0.1083	0.1108	-0.0873	-0.2224	-0.1778	-0.3296	-0.3271
Kurtosis	3.039	2.8597	2.6958	3.1605	3.1527	2.9845	3.4679
Min	-2.9159	-2.8454	-2.8261	-3.9422	-4.0807	-4.7017	-5.2579
Max	3.2082	2.8075	2.674	3.3246	3.3906	2.365	2.8305
T=64,000							
Mean	0.0297	0.0104	0.014	-0.0143	0.0331	0.0217	-0.0157
Median	0.0264	0.0267	0.011	-0.0126	0.0051	0.0339	0.0151
Interquartile	1.2434	1.2844	1.3361	1.4023	1.4085	1.3458	1.324
Std	0.9645	0.9765	0.9941	1.0099	1.015	0.9939	1.0104
Skewness	-0.0331	-0.0179	-0.0239	-0.0377	-0.0444	-0.1118	-0.1477
Kurtosis	3.2684	2.9938	3.0036	3.1543	2.9269	2.8444	3.0043
Min	-3.6638	-2.9635	-3.0832	-3.2335	-3.1705	-3.6821	-3.4006
Max	3.5654	3.37	3.3409	3.5159	2.9356	2.6278	3.3023

Figure 3.21: Distribution of the rescaled sampling error $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$ when states are observed.



Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000

The rescaled sampling error in transition probability parameter p_{11}

Figure 3.21 displays distributions of the rescaled sampling error in p_{11}, which we denote as ξ_{11}. Along the main diagonal it can be observed that the usual asymptotics seems to work well most of the time, except that the mild negative skewness, which arises as a result of transition probability matrix P getting closer to the boundary case, is not washed away. This behaviour is not surprising because this is what the heuristic theory would suggest. As the run length T doubles, the expected number of transitions from *state 1* to *state 1* will more than double by the fixed amount $\frac{g_{22}}{10}$. This is because as we move down the main diagonal, we are simultaneously changing the probability of transiting from *state 1* to *state 1* in a way that it is more likely that one remains in *state 1* given one is currently in *state 1*. Therefore, as T goes to infinity, the information about the parameter p_{11} also goes to infinity, and hence, the distribution of the rescaled sampling error will converge to the standard normal distribution.

Under the usual asymptotic sequence, which is described by each column, the expected number of transitions from *state 1* to *state 1* doubles as the run length T doubles. Consequently, there is more and more information about the transitions from *state 1* to *state 1*, so that the parameter p_{11} is consistently estimated and, therefore, the usual asymptotics work well.

If we move across columns along any row, then T is held fixed whilst P and G vary. This will result in the expected number of transitions from *state 1* to *state 1* to increase by $\frac{g_{22}}{10}$. Subsequently, as G approaches zero additional information about the transition parameter p_{11} is gained. However, as it was mentioned previously, the expected number of transitions from *state 2* to *state 2* will also increase by $\frac{g_{22}}{10}$. As a consequence, once we move to a particular state, we will tend to remain in that state longer, and this will result in some sample realisations providing more information about the transition parameters, and some sample realisations providing less information about the transition parameters. Although the expected occupancy time in either of the regimes remains constant, its standard deviation will be higher the closer the G matrix is to zero for a given T. This will

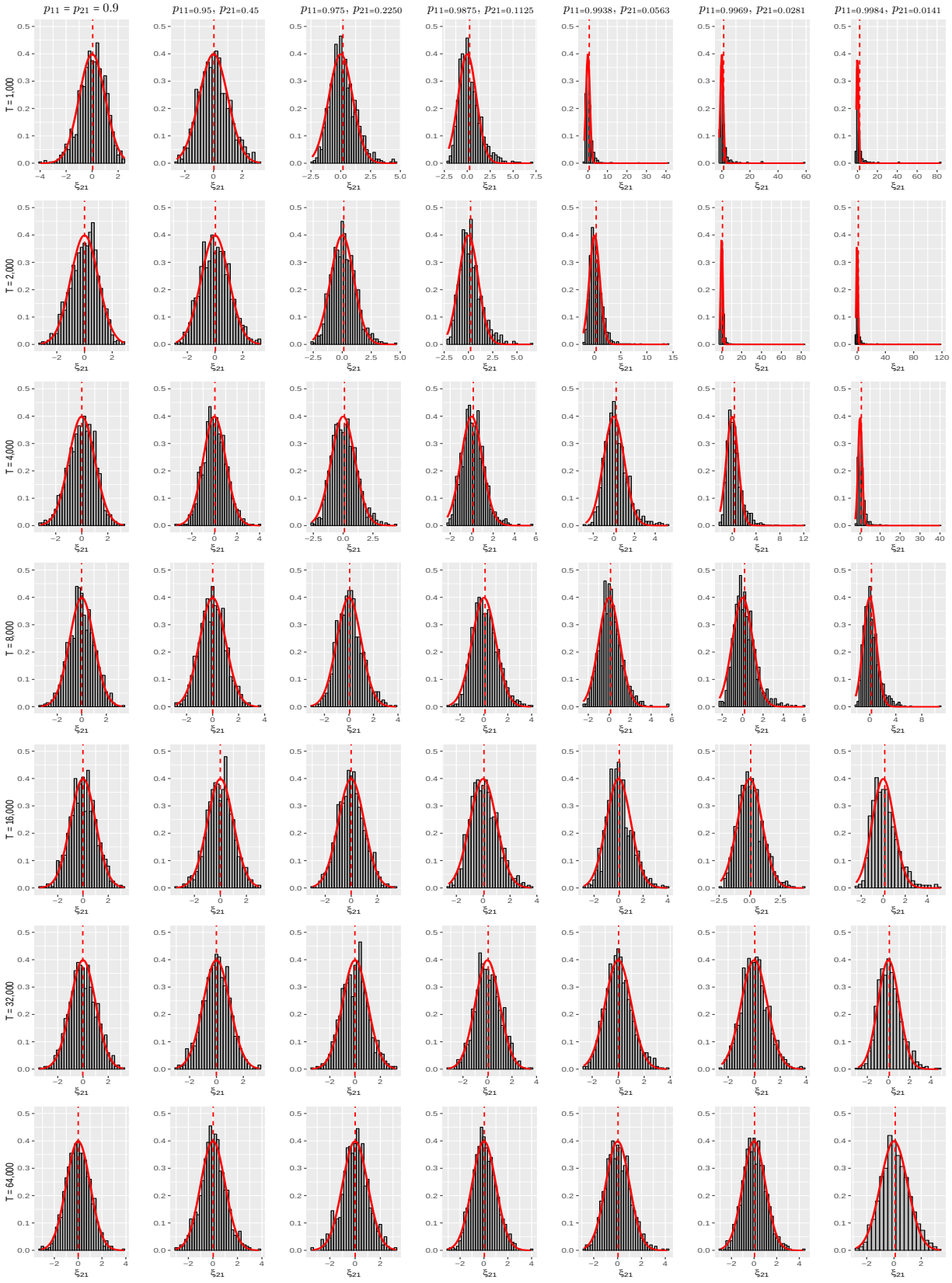
this reason, realisations with no observations from *state 2* have been omitted from the estimation of θ_2. There are six cells in total in the upper right region for which this is the case.

Table 3.20: Summary statistics for the rescaled sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are observed.

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	0.0395	0.0291	0.1894	0.2609	0.6729	1.3187	2.1721
Median	0.0927	0.0197	0.0588	0.0468	0.1369	0.2189	0.2955
Interquartile	1.3154	1.3935	1.3389	1.4534	1.5804	1.504	1.8989
Std	0.9901	1.0252	1.0499	1.1648	2.8143	5.1102	8.2099
Skewness	-0.381	0.2396	0.6573	1.1735	8.3548	7.5207	7.3312
Kurtosis	3.2945	2.8336	3.9122	5.4951	102.5884	74.2075	65.8915
Min	-4.0041	-2.5769	-2.3935	-2.0102	-1.7778	-1.7003	-1.1937
Max	2.4987	3.2647	4.6926	6.9853	40.9402	58.7546	83.6905
T=2,000							
Mean	0.0006	0.0269	0.1321	0.2192	0.3108	0.6999	1.6034
Median	0.0807	-0.0073	0.08	0.0999	0.0761	0.1417	0.2481
Interquartile	1.3395	1.4729	1.387	1.3886	1.3844	1.538	1.7568
Std	0.9896	1.0418	1.0396	1.1268	1.3328	3.5676	6.7098
Skewness	-0.2531	0.2193	0.419	0.9551	3.146	14.7468	9.6887
Kurtosis	2.8246	2.8128	3.4386	4.8047	25.6939	306.5324	130.4943
Min	-3.1419	-2.8262	-2.5457	-2.2	-2.1337	-2.0147	-1.6885
Max	2.8801	3.265	4.6455	6.4721	14.0809	83.1123	118.3859
T=4,000							
Mean	0.0058	0.041	0.1073	0.1515	0.2198	0.3707	0.6069
Median	0.0529	-0.0049	0.0555	0.0778	0.0673	0.1462	0.1579
Interquartile	1.4375	1.3452	1.4456	1.3366	1.3387	1.3724	1.6431
Std	1.0267	1.0092	1.0232	1.0158	1.1022	1.3451	2.2539
Skewness	-0.1887	0.1303	0.4278	0.5796	0.8995	2.2282	7.8922
Kurtosis	2.8875	3.194	3.3352	4.0637	4.5248	13.929	114.8868
Min	-3.1742	-3.3554	-2.6429	-2.2723	-2.7353	-2.0326	-2.3883
Max	3.2242	4.0727	4.4278	5.5376	5.2188	12.1052	40.0698
T=8,000							
Mean	-0.0123	0.0079	0.057	0.0938	0.1247	0.195	0.2611
Median	-0.0465	0.0073	0.0043	-0.0007	0.0482	0.059	0.0426
Interquartile	1.3893	1.3665	1.3392	1.4059	1.2995	1.3479	1.4337
Std	1.0426	0.9917	1.0165	1.0423	1.0402	1.0764	1.2338
Skewness	-0.0193	0.0948	0.3308	0.4087	0.6789	0.9819	1.6273
Kurtosis	2.9307	2.9839	3.1657	3.19	4.4633	5.1428	9.4829
Min	-3.3247	-2.8294	-3.0889	-2.909	-2.4663	-2.258	-2.1665
Max	3.309	3.5551	3.8465	3.9479	5.6071	6.068	10.7513
T=16,000							
Mean	-0.0041	-0.0081	-0.0088	0.0576	0.0562	0.0751	0.1385
Median	-0.0082	-0.0258	-0.0087	-0.0192	-0.027	-0.0075	0.0123
Interquartile	1.4091	1.3402	1.38	1.395	1.2475	1.287	1.4517
Std	1.0371	0.996	1.0093	1.0274	1.0096	1.0016	1.135
Skewness	-0.0268	0.0322	0.1577	0.3856	0.51	0.5031	0.9542
Kurtosis	2.9009	3.0062	2.9383	3.0844	3.491	3.3422	4.4959
Min	-3.4651	-3.3853	-2.9658	-2.5854	-2.8364	-2.3674	-2.454
Max	3.2752	2.9628	3.3413	3.5545	3.9817	4.401	5.2288
T=32,000							
Mean	-0.0163	-0.0056	-0.0175	0.0586	0.0325	0.0411	0.0753
Median	-0.0578	0.015	0.0108	0.0028	-0.0303	-0.0164	-0.0188
Interquartile	1.4112	1.3732	1.3898	1.411	1.3245	1.3299	1.3739
Std	1.0516	0.9939	1.0384	1.0371	1.0168	0.9995	1.0615
Skewness	0.0637	-0.0766	0.1037	0.1257	0.3659	0.3253	0.5873
Kurtosis	2.8272	2.9707	3.0203	2.8559	3.1868	3.1741	3.6853
Min	-3.4657	-2.9962	-3.4553	-3.2396	-2.6533	-2.5065	-2.8725
Max	3.197	3.193	3.2003	3.5318	3.7566	3.774	4.7224
T=64,000							
Mean	0.003	0.0244	0.0039	0.0317	0.0333	0.0346	0.0863
Median	-0.0125	0.0089	0.0203	-0.0209	-0.0222	-0.0081	0.0219
Interquartile	1.3772	1.2626	1.3064	1.3291	1.4256	1.3546	1.3657
Std	1.0051	1.0133	1.0317	0.9902	1.0378	1.0044	1.0336
Skewness	0.0604	0.0488	-0.02	0.1799	0.2835	0.2437	0.2033
Kurtosis	3.0186	3.2372	3.117	3.0333	3.1296	3.2025	2.8657
Min	-3.1155	-3.0404	-3.3822	-3.0386	-2.8231	-2.9453	-2.8124
Max	3.8844	3.7509	3.2513	3.9013	4.1507	4.3426	3.4128

Note: Summary statistics are conditional on spending a positive number of time periods in State 2.

Figure 3.22: Distribution of the sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are observed.[1]



Note: [1] The distribution is conditional on spending a positive number of time periods in State 2. The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000.

result in a strong mixing of distributions, characterised by multiple modes and higher kurtosis. The peak that we observe on the right of the distribution depicted in the top right plot for $T = 1000$ is due to some replications, 179 out of 1000, having zero number of time periods spent in *state 2*. Hence, there cannot be any transitions from *state 1* to *state 2*, and due to the construction of the MLE, this will result in the estimates of p_{11} being equal to one. Hence, we observe a peak at the positive value of $(1 - p_{11})$. The summary statistics in Table 3.19 indicate presence of high kurtosis in the upper right region for $T = 1000$ and G close to zero. In addition, there is a distortion in the form of a negative skewness and strong median bias. Mean bias is never strong in all cells. Interestingly, the standard deviation of 1.02 remains close to the reference distribution.

The rescaled sampling error in transition probability parameter p_{21}

Next we consider the behaviour of the rescaled sampling error in the transition parameter p_{21} denoted as ξ_{21}. In Figure 3.22 the main diagonal displays the sampling behaviour under the alternative asymptotic sequence Case 2. As the run length T doubles, the expected number of transitions from *state 2* to *state 1* remains constant, which is shown in (3.50). This is the result of simultaneous change in the probability transition matrix P towards the boundary case presented in (3.46), so that it is less likely that one will leave *state 2* given one is currently in *state 2*. Hence, the expected number of transitions from *state 2* to *state 1* remains constant, whilst the expected number of transitions from *state 2* to *state 2* is increasing to infinity as T goes to infinity. In other words, under the alternative asymptotic sequence Case 2, the expected number of transitions from *state 2* to *state 1* remains stochastically bounded away from infinity. At this stage it is not clear whether the distribution of the rescaled sampling error in p_{21} to settle down. However, the distribution of the rescaled sampling error is not substantially far from the standard normal distribution, which can be observed by comparing it to the superimposed reference distribution. Furthermore, it seems like the deviation from the standard normal is stable as the run length T gets larger. Therefore, one would expect the supremum of the difference between the rescaled sampling error and the standard normal distribution to remain bounded. From Table 3.20 we also see that the positive skewness does not wash away,

although the mean and the median bias remain reasonably low.

Under the usual asymptotic sequence the expected number of transitions from *state* 2 to *state* 1 doubles as the run length T doubles. The consequence of this is that the information about the transitions from *state* 2 to *state* 1 increases to infinity, so that the parameter p_{21} is consistently estimated. Therefore, the behaviour of the rescaled sampling error ξ_{21} approaches that of a standard normal distribution.

If one moves across columns along any row, one changes P and G whilst holding the run length T constant. This induces a change in the probability of transiting from *state 2* to *state 1*, whilst the steady state probability of being in *state 2* remains constant. As a result, one is less likely to leave *state 2*, given one is in *state 2*. This results in the expected number of transitions from *state 2* to *state 1* to decrease by a half as one moves across the columns. Therefore, the information about the transition parameter p_{21} is decreasing, and the usual asymptotic theory does not seem to work well as G decreases. The closer to zero the G matrix is for a given run length, the worse things start to behave. This is manifested in terms of increasing positive skewness, severe kurtosis, large positive values, strong mean bias and relatively milder median bias.

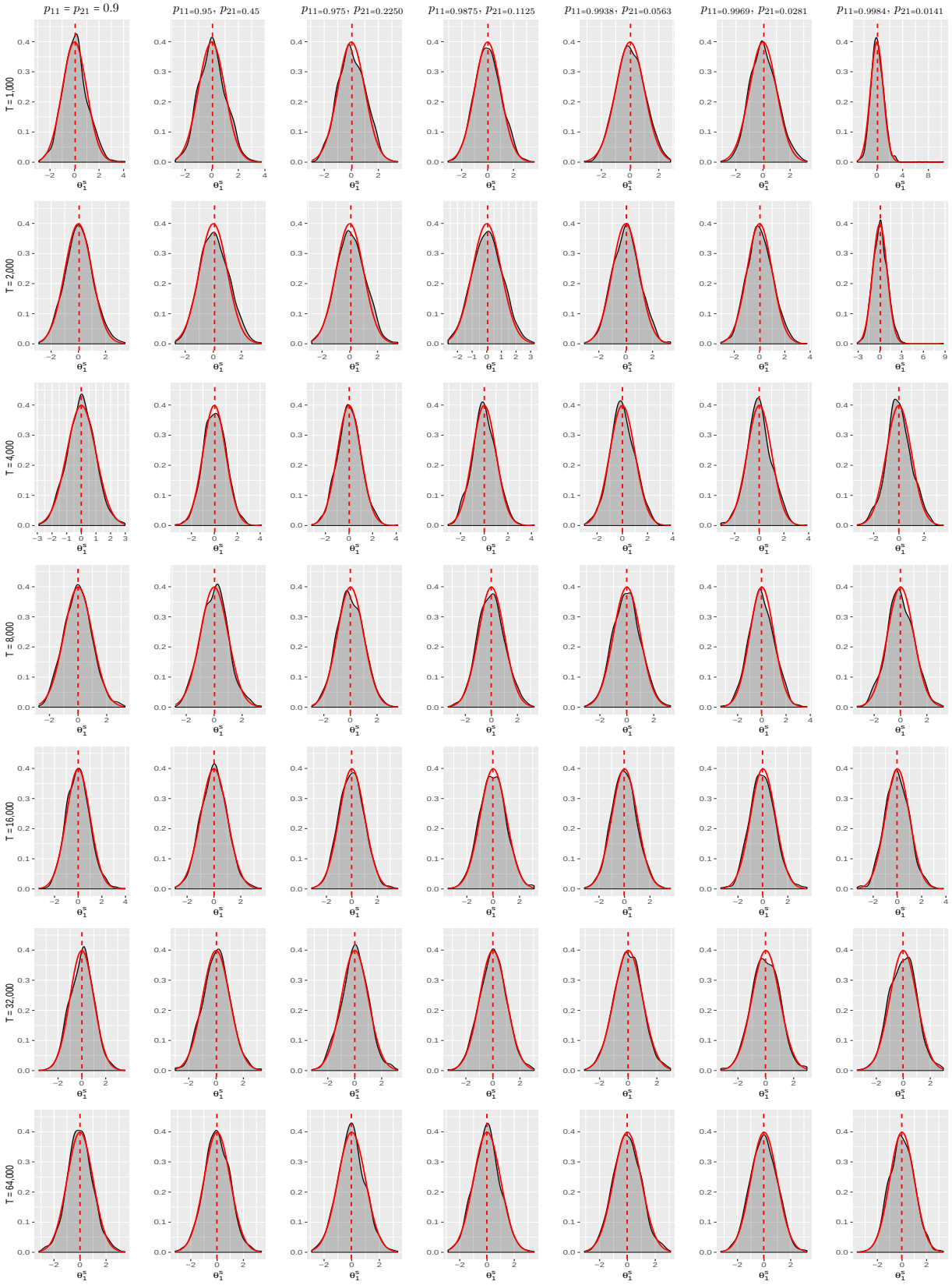
In summing up the empirical results, there seem to be key differences that distinguish the behaviour under the asymptotic sequence Case 2 from the alternative asymptotic sequence Case 1 presented in previous section.

First, is that the expected number of time periods spent in *state 2* increases to infinity as T goes to infinity. Second, the expected number of transitions from one state to a different state remains constant, but the expected number of transitions from one state to the same state increases to infinity. Third, this alternative asymptotic sequence Case 2 pushes the parameter p_{21} to the lower boundary of the parameter space, which is manifested in the positive skewness. However, normal distribution might not work well in the vicinity of the boundary of the parameter space because it assigns a probability mass beyond the boundary.

Table 3.21: Summary statistics for the rescaled sampling error $\theta_1^s = \frac{(\hat{\theta}_\cdot - \theta_\cdot)}{se(\hat{\theta}_\cdot)}$ when states are not observed.

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	0.0519	0.0562	0.0424	0.029	0.0336	0.0607	0.0994
Median	0.0317	0.021	-0.031	0.0155	-0.0085	0.0174	0.0301
Interquartile	1.2925	1.3565	1.4252	1.411	1.3539	1.3969	1.3177
Std	1.0087	1.0008	1.0052	1.0109	0.9901	0.9976	1.0673
Skewness	0.2663	0.18	0.1045	0.1262	0.0563	0.1904	1.5196
Kurtosis	3.3251	2.932	2.6878	2.8936	2.7544	2.8898	14.3547
Min	-2.9023	-2.7396	-2.927	-3.0424	-3.1747	-3.1036	-3.0436
Max	4.1156	3.7358	3.4315	3.5851	2.8237	3.2363	10.3155
T=2,000							
Mean	0.0536	0.084	0.0745	0.0528	0.0507	0.0388	0.0584
Median	0.0411	0.0542	0.0306	0.0589	0.0402	0.0093	0.0487
Interquartile	1.327	1.4438	1.4249	1.4118	1.391	1.3354	1.4151
Std	1.0194	1.0287	1.0368	1.0314	1.0137	1.0198	1.0497
Skewness	0.1326	0.1648	0.0953	0.0825	0.0914	0.117	0.6641
Kurtosis	3.0094	2.7752	2.8788	2.7559	2.9188	2.9679	7.3204
Min	-2.8969	-2.7038	-2.7503	-2.6469	-3.0681	-3.053	-3.12
Max	3.4223	3.4186	3.4656	3.2391	3.3402	3.7438	8.7495
T=4,000							
Mean	-0.0062	0.0354	0.0147	0.0074	0.0049	-0.0092	-0.002
Median	-0.0082	0.0375	0.0041	-0.0225	-0.024	-0.0392	-0.0425
Interquartile	1.2771	1.3895	1.3172	1.354	1.3009	1.2931	1.2525
Std	0.9515	1.0028	1.0043	1.0096	0.9925	1.009	1.006
Skewness	0.0409	0.0279	0.0638	0.1096	0.0547	0.0806	0.0485
Kurtosis	3.0199	3.1734	3.1279	3.076	2.997	3.1533	3.114
Min	-2.8946	-3.4104	-3.1645	-3.0603	-2.9989	-3.1241	-3.307
Max	2.9798	4.109	4.1303	4.2259	3.815	3.8645	3.5271
T=8,000							
Mean	-0.0045	0.0263	-0.0055	-0.0141	-0.0119	-0.0238	-0.0144
Median	-0.015	0.0477	-0.0571	-0.0457	-0.0301	-0.0633	-0.0461
Interquartile	1.33	1.3251	1.3965	1.387	1.3652	1.3752	1.392
Std	0.979	0.9829	1.0026	1.0278	1.0202	1.0414	1.0287
Skewness	0.1792	0.1435	0.1166	0.0963	0.0858	0.0907	0.0621
Kurtosis	3.0096	3.0299	2.8581	2.9531	2.934	3.0076	2.9484
Min	-2.743	-2.7822	-2.9149	-3.3267	-3.1494	-3.3805	-3.287
Max	3.2785	3.3805	3.5652	3.2785	3.2432	3.7203	3.2452
T=16,000							
Mean	-0.0013	0.0117	-0.0201	-0.0269	-0.0258	-0.0287	-0.0285
Median	-0.0302	0.0114	-0.019	-0.0575	-0.0469	-0.0492	-0.0547
Interquartile	1.3677	1.3533	1.3484	1.3485	1.3206	1.3497	1.4034
Std	0.9887	0.977	0.99	1.0117	0.9921	1.0173	1.0086
Skewness	0.246	0.1159	0.1085	0.089	0.1098	0.1816	0.0869
Kurtosis	3.2655	3.0989	3.0547	3.0698	3.0835	3.2787	3.1828
Min	-3.3257	-2.8828	-3.1533	-3.4349	-3.1415	-3.3765	-3.3136
Max	3.9566	3.5397	3.596	3.1137	3.6125	3.5311	3.7855
T=32,000							
Mean	0.0037	0.0221	-0.0007	-0.005	-0.0115	-0.0237	-0.0211
Median	0.05	0.0419	0.0161	0.012	-0.0364	-0.029	-0.003
Interquartile	1.3462	1.2941	1.3087	1.3167	1.3154	1.3824	1.3763
Std	1.0036	0.9959	0.9892	0.9955	0.9774	0.9997	0.9996
Skewness	0.0077	0.027	-0.0023	0.0019	0.0248	-0.0127	-0.0006
Kurtosis	3.0617	3.0742	3.0799	3.0227	2.9721	3.0444	2.9842
Min	-3.6062	-2.9916	-3.1965	-3.321	-3.0534	-3.234	-3.4149
Max	3.6186	3.3179	3.1856	3.0462	2.952	2.9526	2.9637
T=64,000							
Mean	-0.0031	-0.0122	-0.0268	-0.0133	-0.0196	-0.0243	-0.02
Median	-0.0282	-0.0209	-0.0441	-0.0098	-0.0353	-0.0296	-0.0103
Interquartile	1.2681	1.3326	1.2766	1.2905	1.3591	1.3869	1.3597
Std	0.9843	0.9837	0.9845	0.9821	0.9925	1.0031	1.005
Skewness	0.0846	0.0454	0.1012	0.0989	0.096	0.028	0.0147
Kurtosis	3.2744	3.1763	3.0878	3.0474	2.9441	2.9381	3.0143
Min	-3.1642	-3.1755	-2.897	-2.8592	-3.0852	-3.25	-3.6424
Max	3.4553	3.3666	3.3315	3.4566	3.0984	3.2226	3.3576

Figure 3.23: Distribution of the rescaled sampling error of $\theta_1^s = (\hat{\theta}_1 - \theta_1)/se(\hat{\theta}_1)$ when states are not observed.



Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000

Unobserved states

Next we consider a world in which the states cannot be observed. In previous section it was confirmed that not being able to observe the states adds an additional complication to the estimation problem. Therefore, we expect to observe, to some degree, similar consequences of not being able to observe the states. However, it will be interesting to see what will be the nature and the magnitude of the differences under the alternative asymptotic sequence Case 2.

The rescaled sampling error in copula parameter θ_1 in state 1

In Figure 3.23 along the main diagonal we can observe that the behaviour of the rescaled sampling error under the alternative asymptotic sequence Case 2 is not substantially affected by the lack of observability. Unsurprisingly, the behaviour under the usual asymptotics seems also to be unaffected. This is because as the run length T increases, the amount of information about the copula parameter θ_1 in *state 1* increases to infinity, and the rescaled sampling error becomes better and better approximated by the standard normal distribution. Moreover, we are not getting close to the boundary of the parameter space because the transition probability matrix P is being held constant.

We could also analyse the impact of not being able to observe the states on the behaviour of the rescaled sampling error as G gets very close to zero. From the plots it can be observed that the behaviour of the rescaled sampling error does not deteriorate rapidly, and G has to get very close to zero in order for the distortions to become noticeable. The summary statistics in Table 3.26 indicate that these distortions appear in the form of high positive skewness and stronger kurtosis. In addition, there is a sizeable mean bias, although absence of strong median bias. These distortions can be observed in the upper right region of Figure 3.23. Also, in this region the expected number of time periods spent in *state 1* has higher variability. In some realisations we would spend a modest number of time periods in *state 1*, and in some realisations we will spend most of the time in *state 1*. When we spend all of the time in *state 1* this implies that none of the time will be spent in *state 2*. This will result, in contrast to the observed case, in severe distortions in the behaviour of the

rescaled sampling error in θ_2, which will "contaminate" the distribution of the rescaled sampling error in θ_1 due to the way the likelihood function is constructed. Furthermore, it is suspected that potentially all of the parameters will be "contaminated".

These results indicate that the lack of observability does not seem to affect substantially the estimates of θ_1. One of the most notable feature of the alternative asymptotic sequence Case 2 is that, nowhere in the plots the expected number of time periods spent in either state decreases below its initial level, that is the expected number of time periods in the first cell at $T = 1000$. This could potentially explain why the lack of observability does not create severe distortions in the behaviour of the rescaled sampling error in θ_1 compared to the alternative asymptotic sequence Case 1.

The rescaled sampling error in copula parameter θ_2 in state 2

Next we consider the distribution of the rescaled sampling error in θ_2. Figure 3.24 displays the sampling distribution of θ_2^s in *state 2*.

The behaviour along the main diagonal does not seem to be substantially affected by the lack of observability. At the run length $T = 1000$, there seem to be some minor distortions such as considerable mean bias, notable negative skewness, and slight kurtosis. However, it can be observed that as the run length T increases these distortions gradually wash away, and the rescaled sampling error begins to be well approximated by the reference distribution. This is because under this alternative asymptotic sequence, the number of time periods spent in *state 2* doubles when the run length T doubles, and the expected number of time periods spent in *state 2* increases to infinity as the run length T increases to infinity. The copula parameter θ_2 from *state 2* is being well determined, and the usual asymptotic theory works reasonably well in most of the plots down the main diagonal.

If one goes vertically down the column, the behaviour of the rescaled sampling error is starting to be better and better approximated by the standard normal distribution. We can notice that the mean bias, high kurtosis, and a negative skewness present in the first cell for $T = 1000$ are gradually washed away with larger T.

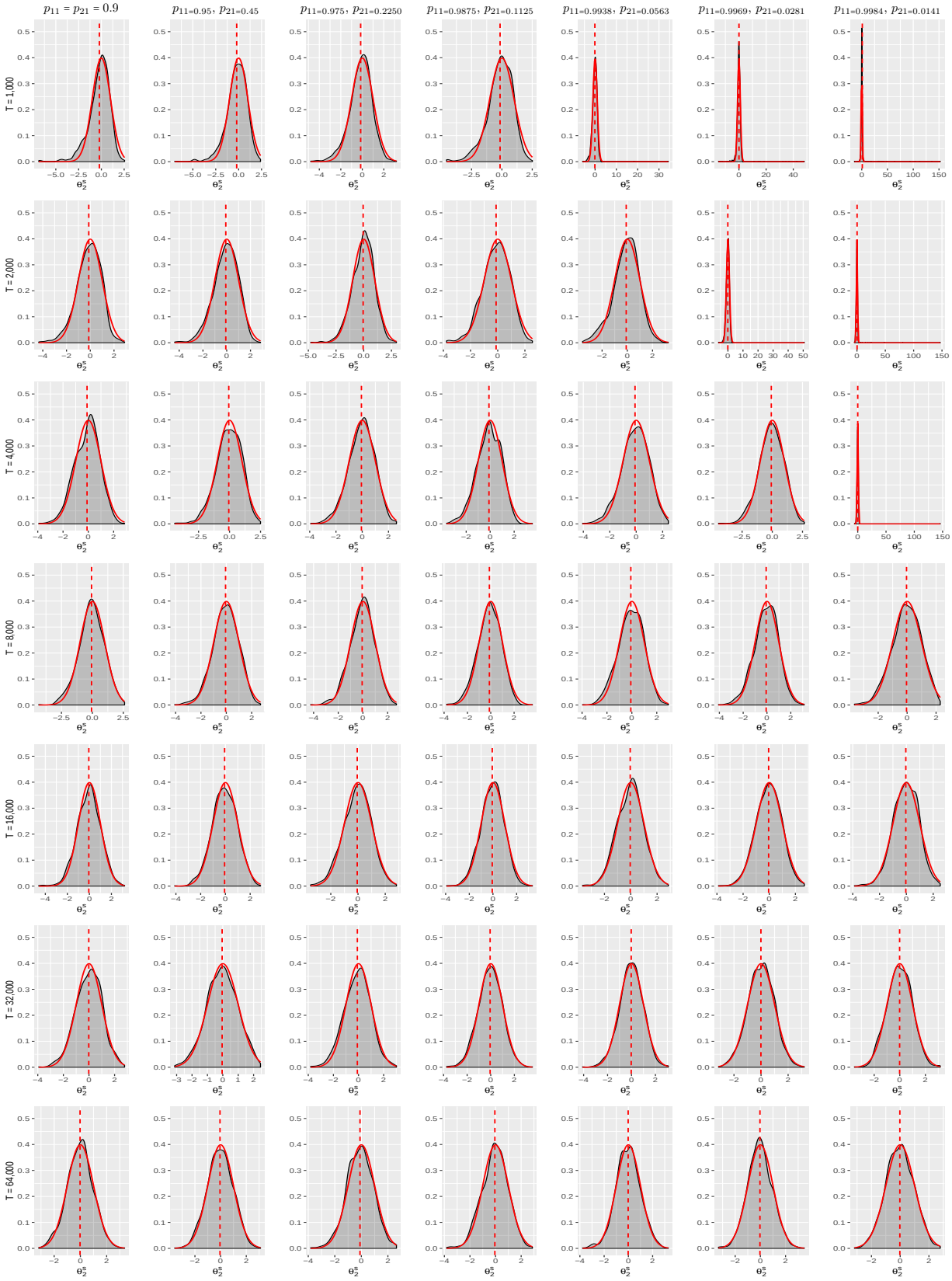
We can also analyse how things behave as we keep the run length T constant and let matrix G get very close to zero. As we move across the columns along each row in Figure

Table 3.22: Summary statistics for the rescaled sampling error $\theta_2^s = \frac{(\hat{\theta}_\cdot - \theta_\cdot)}{se(\hat{\theta}_\cdot)}$ when states are not observed.

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	-0.2236	-0.2134	-0.1542	-0.1505	-0.1081	0.0358	1.0256
Median	-0.0497	-0.0935	-0.0761	-0.0392	-0.0476	-0.0243	0.0000
Interquartile	1.3468	1.4083	1.2815	1.3445	1.3504	1.196	1.0816
Std	1.1552	1.1106	1.0463	1.0393	1.6286	3.2679	9.6125
Skewness	-1.2151	-0.9723	-0.5451	-0.7754	10.7641	8.9906	11.1605
Kurtosis	6.4128	5.5525	4.0212	4.2621	219.6702	117.414	144.4863
Min	-6.8697	-6.9778	-4.8038	-4.5557	-5.8852	-15.0514	-14.2115
Max	2.546	2.3882	3.1749	2.5136	34.4135	48.1079	152.4146
T=2,000							
Mean	-0.1276	-0.0772	-0.0532	-0.1201	-0.0844	-0.0669	0.2421
Median	-0.0296	-0.0007	0.0278	-0.06	-0.015	-0.0097	-0.0505
Interquartile	1.3911	1.3722	1.3347	1.3766	1.2848	1.3186	1.3932
Std	1.0352	1.0382	0.9914	1.0096	1.0261	1.9109	5.7849
Skewness	-0.5243	-0.5099	-0.4543	-0.3933	-0.4173	18.6735	19.7766
Kurtosis	3.6559	3.4675	3.8368	3.2781	3.3389	500.3111	456.8184
Min	-4.3375	-4.3751	-4.7899	-3.7606	-3.5482	-6.2583	-4.4325
Max	2.8894	2.8433	2.9538	2.5513	3.2549	50.7179	146.8918
T=4,000							
Mean	-0.0983	-0.0615	-0.0596	-0.1007	-0.0885	-0.0811	0.0456
Median	-0.0041	-0.0116	0.0073	-0.0578	-0.0423	-0.0406	-0.0435
Interquartile	1.3981	1.3834	1.3505	1.468	1.3534	1.3836	1.3781
Std	0.9927	1.0273	0.9965	1.0164	1.0319	1.0243	4.7327
Skewness	-0.304	-0.4685	-0.3398	-0.3185	-0.3599	-0.329	29.2494
Kurtosis	3.0296	3.4236	3.1763	2.9948	3.0992	3.285	901.6457
Min	-3.9192	-4.4226	-3.9267	-3.6319	-4.0577	-4.4529	-5.8015
Max	2.8681	2.4978	2.607	3.5215	2.431	2.6478	145.8088
T=8,000							
Mean	-0.0652	-0.0582	-0.0603	-0.1162	-0.082	-0.0677	-0.0393
Median	-0.0582	-0.0182	-0.0057	-0.0753	-0.0415	-0.0089	0.0068
Interquartile	1.3428	1.3581	1.274	1.36	1.413	1.3541	1.3453
Std	1.0227	1.0088	0.9993	1.0029	1.0258	1.019	0.995
Skewness	-0.2254	-0.3289	-0.3002	-0.2391	-0.2401	-0.3142	-0.3153
Kurtosis	3.1067	3.2088	3.221	3.073	2.8969	3.2169	2.9539
Min	-4.332	-4.0683	-4.2374	-3.7841	-3.9605	-4.0735	-3.6412
Max	2.6046	2.6965	2.7203	3.5689	2.9308	3.1357	2.2889
T=16,000							
Mean	-0.0513	-0.0646	-0.0574	-0.0614	-0.0414	-0.0476	-0.0402
Median	-0.0096	-0.0519	-0.0041	-0.0147	-0.0058	-0.0254	-0.0055
Interquartile	1.4126	1.3857	1.4016	1.2958	1.3394	1.3244	1.3456
Std	1.0713	0.9961	1.0163	0.982	0.9896	0.9747	0.9446
Skewness	-0.225	-0.1425	-0.2395	-0.1829	-0.1162	-0.1715	-0.2696
Kurtosis	3.4506	2.9504	3.0453	3.1743	3.013	3.0797	3.2704
Min	-4.5942	-4.0956	-3.4779	-3.9179	-3.662	-3.8821	-3.8722
Max	3.194	2.8369	2.8237	3.3344	2.8541	2.6788	2.5253
T=32,000							
Mean	-0.0166	-0.0445	-0.0917	-0.0521	-0.0241	-0.0084	-0.0138
Median	0.0527	-0.0399	-0.0374	-0.028	-0.0054	0.0246	-0.0163
Interquartile	1.4155	1.3866	1.4077	1.3388	1.2873	1.3061	1.3313
Std	1.024	0.9992	1.0482	1.0198	0.9861	0.9593	0.9664
Skewness	-0.2119	-0.1099	-0.151	-0.1094	-0.1772	-0.0781	-0.0003
Kurtosis	3.0391	2.7856	3.1057	3.1656	3.314	2.9657	2.9409
Min	-3.9216	-3.1235	-3.809	-3.7391	-4.0901	-3.1434	-3.54
Max	2.7802	2.4493	3.0179	3.5289	3.081	3.1956	3.1559
T=64,000							
Mean	-0.0261	-0.0469	-0.1015	-0.054	-0.0176	-0.0323	-0.0227
Median	-0.0075	-0.0497	-0.0762	-0.0232	-0.0038	-0.0572	-0.0062
Interquartile	1.2835	1.3308	1.3709	1.3203	1.3113	1.2851	1.3353
Std	0.9851	0.9943	0.9954	1.0099	1.012	0.9738	0.9625
Skewness	-0.0485	-0.0736	-0.1162	-0.2292	-0.0985	0.0565	0.0079
Kurtosis	3.0462	2.8721	3.1561	3.1451	3.3001	3.1705	2.7938
Min	-3.1698	-3.5756	-3.8194	-3.8242	-3.7815	-3.3252	-3.1829
Max	3.3653	3.1096	2.6458	2.8785	3.2984	3.4873	2.7864

Note: Summary statistics are conditional on spending a positive number of time periods in *state 2*.

Figure 3.24: Distribution of the rescaled sampling error of $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$ when states are not observed.



Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000

3.24, we can observe the lack of observability results in similar set of distortions as in the observed case. However, summary statistics in Table 3.22 indicate that the closer the G matrix is to zero, the stronger is the mean bias and kurtosis, accompanied by a higher standard deviation and larger positive values.

The rescaled sampling error in transition probability parameter p_{11}

Figure 3.25 displays distributions of the rescaled sampling error in p_{11} when states cannot be observed, which we denote as ξ_{11}. The behaviour under the alternative asymptotic sequence along the main diagonal is not substantially affected by the lack of observability, and the usual asymptotic theory seems to work well most of the time. However, as a result of the transition probability matrix P getting closer to the boundary case, a mild negative skewness is present and does not seem to wash away with larger T.

Under the usual asymptotic sequence, which is displayed by each column, the parameter p_{11} is consistently estimated. The negative skewness is gradually washed away with increasing T and, therefore, the usual asymptotics work well.

If we move across the columns along any row, the behaviour starts to deteriorate more rapidly, as opposed to the observed case. Consequently, as G gets very close to zero severe distortions begin to emerge. These are manifested in the form of strong mean bias, increased standard deviation, deep negative skewness, strong kurtosis, and large negative values. The median bias does not seem to be strong in most of the scenarios.

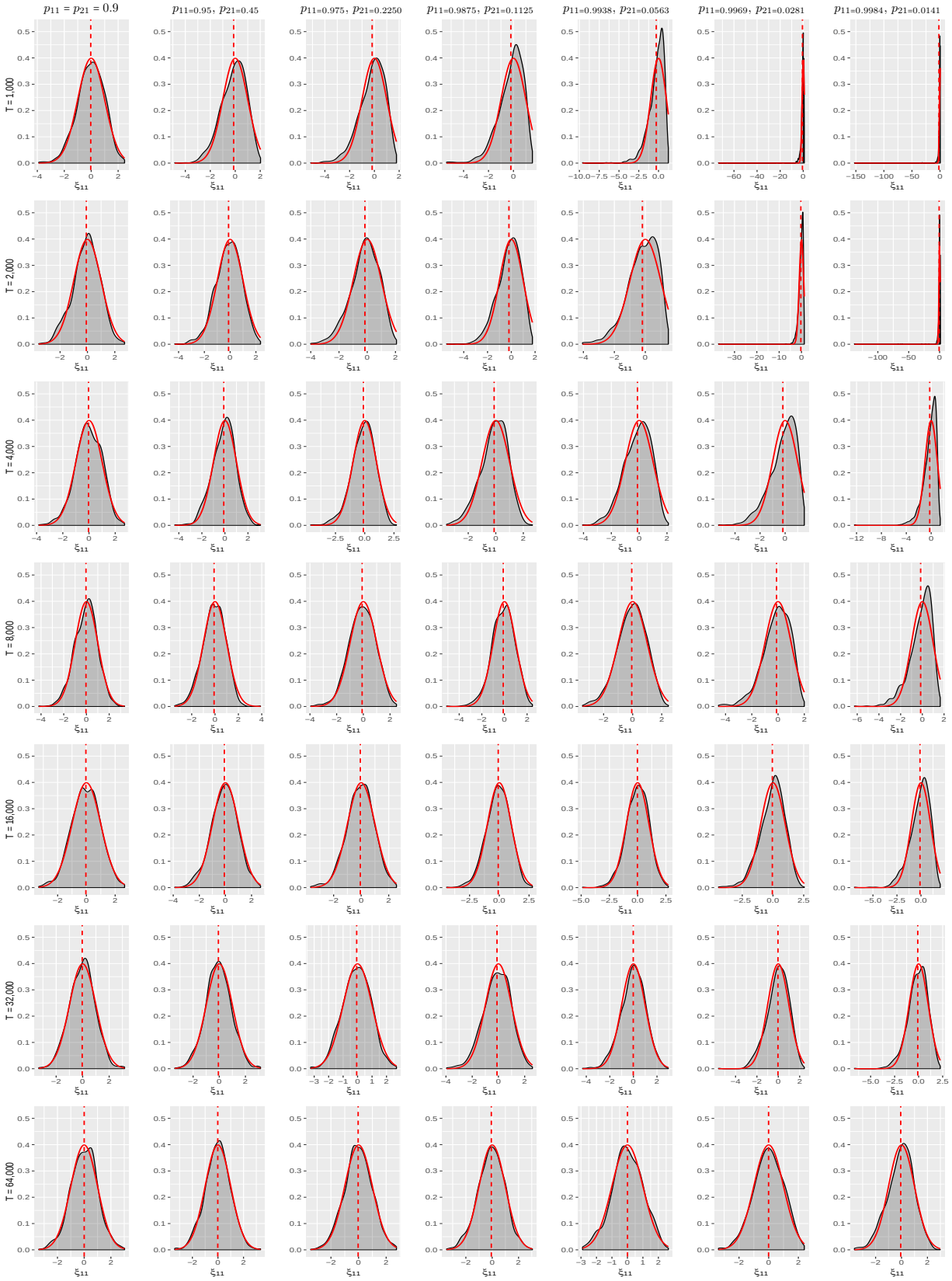
The rescaled sampling error in transition probability parameter p_{21}

Next we consider the behaviour of the rescaled sampling error in the transition parameter p_{21}, which we denote as ξ_{21}. In Figure 3.26 the main diagonal displays the sampling behaviour under the alternative asymptotic sequence Case 2. It can be observed that the lack of observability does not affect substantially the behaviour of the rescaled sampling error, except a noticeable distortion in the first plot for $T = 1000$. This distortion comes in the form of strong mean bias, deep skewness, high kurtosis, and negative values of considerable size. As T increases, these distortions wash away except for the skewness, which is now positive due to the matrix P approaching the boundary case as in (3.46).

Table 3.23: Summary statistics for the rescaled sampling error $\xi_{11} = \frac{(\hat{p}_{\cdot\cdot} - p_{\cdot\cdot})}{se(\hat{p}_{\cdot\cdot})}$ when states are not observed.

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	-0.0339	-0.1289	-0.1821	-0.2084	-0.2888	-0.6596	-1.9039
Median	0.0159	-0.0191	-0.0422	-0.0078	-0.0434	-0.1961	-0.0327
Interquartile	1.3623	1.41	1.4108	1.2894	1.1941	1.3834	1.1786
Std	0.995	1.0132	1.0367	1.0106	1.0435	3.4582	10.7911
Skewness	-0.341	-0.5574	-0.8131	-1.1304	-1.9835	-16.5913	-10.5307
Kurtosis	3.1718	3.2812	3.8803	5.1382	11.8552	326.1477	123.2791
Min	-3.8917	-4.8469	-5.1487	-5.3621	-9.6015	-73.4215	-152.5791
Max	2.4786	2.0229	1.7821	1.5207	1.2394	0.9264	0.4419
T=2,000							
Mean	-0.0796	-0.1219	-0.167	-0.2002	-0.1887	-0.2783	-0.747
Median	-0.0259	-0.0702	-0.0781	-0.0798	-0.0697	0.0306	-0.1402
Interquartile	1.2924	1.3278	1.352	1.3502	1.3469	1.323	1.2349
Std	0.9917	0.9898	1.0152	0.9992	0.9706	1.5514	5.7968
Skewness	-0.2652	-0.3576	-0.5256	-0.7153	-0.831	-13.8216	-17.3225
Kurtosis	3.0192	3.1889	3.1942	3.8115	3.7096	321.292	356.9663
Min	-3.5577	-4.3003	-4.0062	-5.4423	-4.0543	-37.1658	-137.9045
Max	2.7155	2.3567	2.069	1.7741	1.4875	1.254	1.0257
T=4,000							
Mean	-0.0349	-0.0984	-0.1137	-0.1194	-0.1221	-0.1811	-0.2333
Median	-0.0154	-0.0137	-0.0375	-0.0248	-0.0127	0.041	0.083
Interquartile	1.3981	1.3319	1.3609	1.3187	1.4136	1.4268	1.3238
Std	1.0051	1.0002	1.0217	0.9904	0.9862	1.061	1.1113
Skewness	-0.2775	-0.3399	-0.5308	-0.4983	-0.5247	-0.9572	-2.3476
Kurtosis	3.1185	3.2387	3.5913	3.1077	3.0145	3.892	17.9237
Min	-3.8332	-4.3582	-4.619	-3.5705	-4.0676	-5.3192	-12.3392
Max	2.7144	3.0899	2.7111	2.6592	2.0975	1.53	1.455
T=8,000							
Mean	-0.0422	-0.0822	-0.0897	-0.1041	-0.0627	-0.1142	-0.1556
Median	0.0188	-0.0524	-0.0647	-0.0515	0.0036	0.0184	0.0602
Interquartile	1.3858	1.313	1.3528	1.3838	1.3537	1.4268	1.2842
Std	1.0022	0.977	0.9907	1.012	0.9913	1.057	1.0449
Skewness	-0.2265	-0.2303	-0.269	-0.3932	-0.3869	-0.8095	-1.163
Kurtosis	3.1854	3.1025	3.1239	3.5195	3.0225	3.9478	5.0137
Min	-4.1972	-3.4538	-3.9472	-5.111	-3.4502	-4.5372	-6.2562
Max	3.3559	3.8965	2.4817	2.4698	2.4559	1.9989	1.6488
T=16,000							
Mean	-0.0358	-0.0848	-0.06	-0.0894	-0.0208	-0.0495	-0.1019
Median	-0.0334	-0.0351	-0.0261	-0.0545	0.0266	0.077	0.079
Interquartile	1.3644	1.3642	1.2838	1.3562	1.3622	1.3267	1.365
Std	0.9911	1.0113	0.9943	1.012	0.9958	1.0144	1.0543
Skewness	-0.1414	-0.2037	-0.2748	-0.3225	-0.3679	-0.6426	-1.0382
Kurtosis	2.9962	3.0774	3.2534	3.3009	3.4885	3.7067	5.346
Min	-3.3132	-3.8849	-3.6678	-4.4295	-4.8757	-4.4314	-6.9244
Max	2.6554	2.6877	2.5626	2.7828	2.7155	2.5487	2.0026
T=32,000							
Mean	-0.0338	-0.0638	-0.0587	-0.1092	-0.0131	-0.0299	-0.0923
Median	-0.0049	-0.0708	-0.0303	-0.0574	0.0408	0.0447	-0.0035
Interquartile	1.2892	1.2724	1.3402	1.4096	1.3559	1.3501	1.3639
Std	0.9693	0.9821	0.969	1.0299	1.0385	1.0218	1.0407
Skewness	-0.0352	0.0084	-0.1082	-0.352	-0.3283	-0.5418	-0.7561
Kurtosis	3.1686	3.1709	2.87	3.1571	3.3073	3.5626	4.6477
Min	-3.3123	-3.4008	-3.2421	-3.951	-4.316	-5.5968	-6.6865
Max	3.1646	3.1511	2.6879	2.5878	2.9641	2.4345	2.2469
T=64,000							
Mean	-0.0001	-0.0129	-0.0293	-0.0651	0.0043	-0.006	-0.0601
Median	0.0228	0.0405	-0.0351	-0.0448	0.0045	-0.0009	0.0356
Interquartile	1.3122	1.2853	1.2959	1.3605	1.3312	1.3941	1.3506
Std	0.999	0.9892	1.002	1.0162	1.0037	1.0009	1.0235
Skewness	-0.0914	-0.0864	-0.0789	-0.1484	-0.1511	-0.2294	-0.3749
Kurtosis	3.1371	3.2002	3.0025	3.0644	2.8262	2.8734	3.1441
Min	-3.3984	-3.3801	-3.47	-3.4604	-2.8991	-3.3799	-3.6325
Max	3.0249	3.3295	2.7462	3.0531	2.644	2.399	2.9874

Figure 3.25: Distribution of the sampling error $\xi_{11} = {}^{(\hat{p}_{11} - p_{11})}/_{se(\hat{p}_{11})}$ when states are not observed.

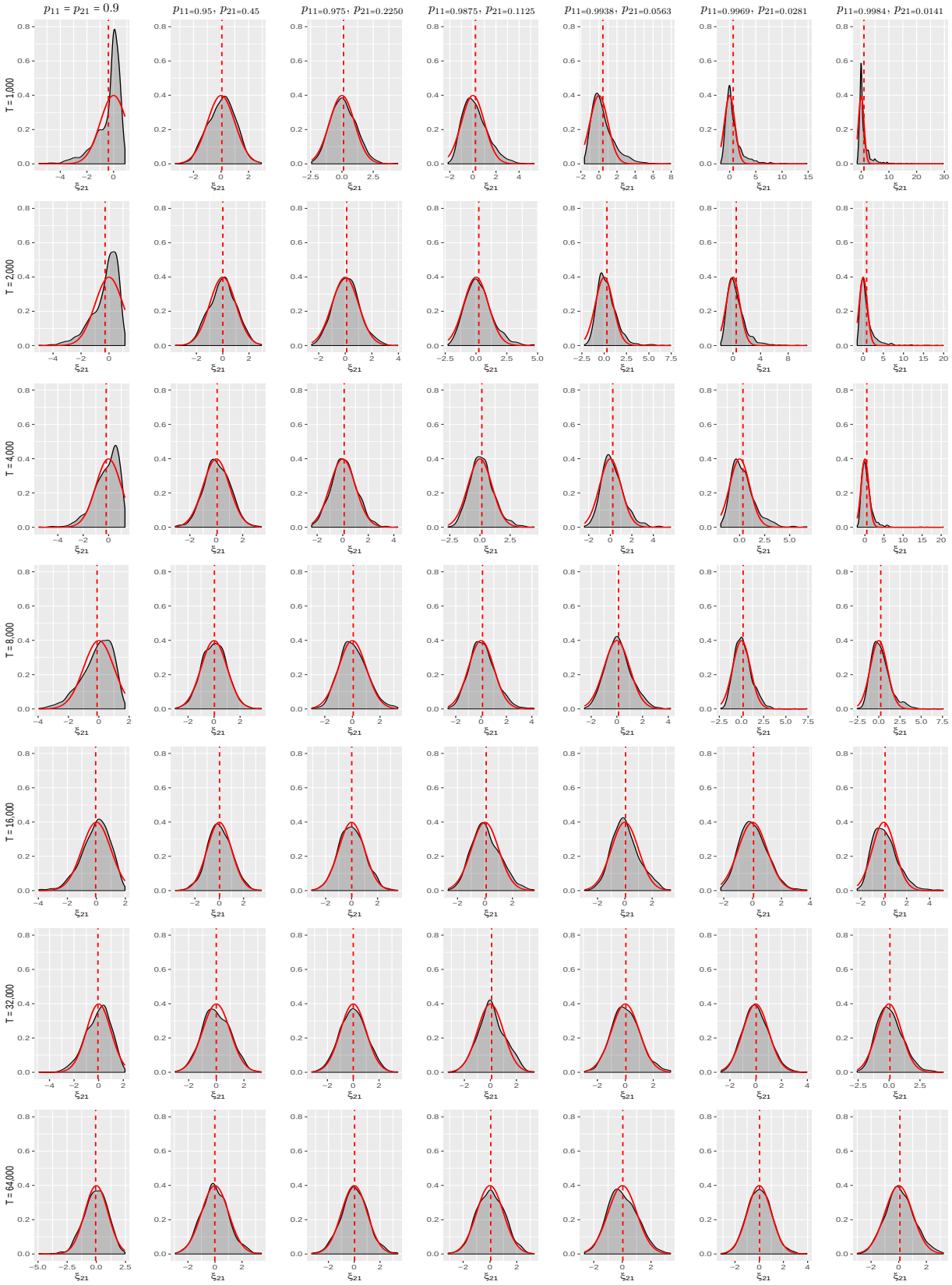


Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000

Table 3.24: Summary statistics for the rescaled sampling error $\xi_{21} = \frac{(\hat{p}_{\cdot\cdot} - p_{\cdot\cdot})}{se(\hat{p}_{\cdot\cdot})}$ when states are not observed.

	p_{11} = 0.9 p_{21} = 0.9	p_{11} = 0.95 p_{21} = 0.45	p_{11} = 0.975 p_{21} = 0.225	p_{11} = 0.9875 p_{21} = 0.1125	p_{11} = 0.9938 p_{21} = 0.0563	p_{11} = 0.9969 p_{21} = 0.0281	p_{11} = 0.9984 p_{21} = 0.0141
T=1,000							
Mean	-0.3849	0.0569	0.1253	0.2241	0.4395	0.6824	0.9905
Median	0.0000	0.1194	0.0722	0.0747	0.1632	0.1866	0.1064
Interquartile	1.0495	1.3731	1.4417	1.4426	1.4966	1.4423	1.4414
Std	0.9678	0.9893	1.0209	1.0862	1.2639	1.6465	2.3835
Skewness	-1.773	-0.1339	0.3063	0.7704	1.3436	2.4741	4.0862
Kurtosis	6.7263	2.8034	3.157	3.6489	5.5325	13.042	31.8631
Min	-5.6082	-3.3261	-2.4182	-2.1496	-1.5984	-1.6765	-1.4301
Max	0.8599	2.9302	4.4582	5.3057	7.9245	14.764	29.6176
T=2,000							
Mean	-0.2797	0.006	0.1047	0.2274	0.3039	0.4695	0.8527
Median	0.0000	0.035	0.0792	0.1245	0.1513	0.1852	0.2505
Interquartile	1.1991	1.3914	1.2993	1.3658	1.306	1.493	1.6079
Std	0.9533	1.0188	0.9743	1.0793	1.1223	1.3662	2.1783
Skewness	-1.2745	-0.0529	0.2569	0.6677	1.392	1.9997	3.6697
Kurtosis	4.7255	2.8316	3.1041	3.7577	7.2337	10.4834	23.0977
Min	-5.0138	-3.6094	-2.5252	-2.2678	-2.2092	-1.7638	-1.4872
Max	1.1415	2.9387	3.9478	4.718	7.4231	10.7146	19.9154
T=4,000							
Mean	-0.1827	0.0631	0.1145	0.1566	0.2327	0.3326	0.485
Median	0.0081	0.0234	0.0849	0.0869	0.0958	0.1252	0.1597
Interquartile	1.3534	1.3302	1.2742	1.2172	1.305	1.3963	1.5268
Std	0.986	0.9949	0.9894	1.0006	1.057	1.1801	1.6763
Skewness	-1.0931	0.0385	0.2419	0.5255	0.8152	1.2147	4.4768
Kurtosis	4.6757	3.0831	3.178	3.6604	4.4609	5.5571	41.3229
Min	-5.4894	-3.1055	-2.4272	-2.6426	-2.4018	-1.8763	-2.0245
Max	1.2968	3.4173	4.3088	4.4904	5.5975	6.7211	20.5657
T=8,000							
Mean	-0.1138	0.0257	0.0623	0.1089	0.1275	0.1799	0.2336
Median	0.0675	0.0195	0.0072	0.0314	0.0535	0.0764	0.0509
Interquartile	1.3973	1.3657	1.3405	1.3338	1.2633	1.2582	1.3628
Std	1.0174	0.9738	1.0092	1.0253	1.0044	1.0242	1.1616
Skewness	-0.7951	0.078	0.3	0.4437	0.422	1.0048	1.2327
Kurtosis	3.3273	3.099	3.0964	3.4339	3.3407	6.2366	6.1014
Min	-3.947	-2.9831	-2.9465	-2.6065	-2.5219	-2.3264	-2.5353
Max	1.7258	3.6394	3.2874	4.1816	4.1755	7.3735	7.6038
T=16,000							
Mean	-0.046	0.0221	0.011	0.0837	0.0483	0.0618	0.1182
Median	0.0657	-0.0148	-0.0131	-0.0174	-0.0534	-0.0173	-0.0206
Interquartile	1.3053	1.3841	1.419	1.4219	1.3122	1.3408	1.467
Std	0.9825	1.0113	1.016	1.0432	1.0011	0.9722	1.1073
Skewness	-0.6589	0.0268	0.1405	0.3848	0.3797	0.4438	0.8423
Kurtosis	3.556	3.1245	2.9762	2.957	3.0377	3.2265	4.1902
Min	-3.9647	-3.6126	-3.0755	-2.6906	-2.9613	-2.2061	-2.3116
Max	1.975	3.4555	3.5565	3.5824	3.348	3.7694	5.1955
T=32,000							
Mean	-0.0579	-0.0017	0.0022	0.0943	0.0387	0.0329	0.0691
Median	0.0588	-0.0552	0.0014	0.0407	-0.0244	-0.0005	-0.0167
Interquartile	1.4026	1.4527	1.4481	1.4037	1.3758	1.3593	1.3789
Std	1.0288	1.0199	1.0382	1.0424	1.0181	0.9977	1.0438
Skewness	-0.5151	0.0806	0.0311	0.131	0.248	0.2744	0.4935
Kurtosis	3.1665	2.7391	2.8494	2.8236	3.0725	3.0205	3.3554
Min	-4.8628	-2.9666	-3.2233	-3.2495	-2.9247	-2.7005	-2.5706
Max	2.1404	3.2585	3.4512	3.3576	3.2366	3.9689	4.3857
T=64,000							
Mean	-0.0547	-0.0099	0.0234	0.0549	0.0296	0.0278	0.0841
Median	-0.0154	-0.0421	0.0163	0.0474	-0.0451	0.0201	0.0253
Interquartile	1.3847	1.347	1.3416	1.4567	1.4391	1.4016	1.3653
Std	1.0146	1.0053	1.038	1.0218	1.0245	1.0018	1.0322
Skewness	-0.3746	0.1632	0.0152	0.049	0.2708	0.1678	0.1847
Kurtosis	3.3018	3.0944	3.1649	2.8102	2.8063	3.0691	2.9042
Min	-4.9195	-2.9295	-3.298	-3.1249	-2.6533	-3.1707	-2.9994
Max	2.4577	3.4271	3.3687	3.2953	3.367	3.9431	3.2219

Figure 3.26: Distribution of the rescaled sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are not observed.



Note: The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000

At this point, one important feature deserves attention in the estimation procedure when states are not observed. When states cannot be observed, the likelihood is constructed in such a way that both p_{21} and p_{22} enter the optimisation procedure. Consequently, it is expected that because the expected number of transitions from *state 2* to *state 2* is going to infinity, p_{22} will be consistently estimated. Hence, due to the sum of probabilities being equal to one, $p_{21} + p_{22} = 1$, p_{21} will also be consistently estimated. In this case, we expect the distribution of the sampling error in p_{21} to be better and better approximated by the standard normal distribution, although a mild positive skewness is expected to remain.

If we move vertically down the columns, the information about the transitions from *state 2* to *state 1* increases to infinity, and the behaviour of the rescaled sampling error ξ_{21} approaches that of a standard normal distribution.

If we move across the columns along any row, the behaviour improves in the second column, and then starts to deteriorate. The improvement in the behaviour is not entirely surprising. If one moves across columns along any row, the P matrix is changing towards the boundary case as in (3.46). In the second cell the P matrix is such that the parameter $p_{21} = 0.45$ has moved away from the upper boundary. Consequently, as G gets very close to zero severe distortions begin to emerge, where $p_{21} = 0.0141$ is now close to its lower boundary. These distortions are manifested in the form of increased standard deviation, deep positive skewness, strong kurtosis, large positive values, strong mean bias and relatively milder median bias. In addition, another possible cause of distortions is that the change in P towards the boundary case results in both the expected number of transitions from *state 1* to *state 1*, and from *state 2* to *state 2* to increase. As it was mentioned previously, the consequence of this is that, once we move to a particular state, we will tend to remain in that state longer. This will result in some sample realisations providing more observations from *state 2*, and some sample realisations providing less observations from *state 2*. Those sample realisations with less observations from *state 2* will provide less information about the copula parameter θ_2, thus creating distortions that could potentially "contaminate" not only estimates of p_{21}, but also all other parameter estimates.

3.4 Empirical application

In this section, we provide an example of applying the alternative asymptotic framework considered in previous sections. We use data from Section 2.5, which comprises S&P500 and FTSE100 indices for the period February 1, 1990 to November 18, 2014, amounting to 1299 observations. We fit to the data the regime-switching copula model introduced in (3.37), and obtain its parameter estimates, which are presented in Table 3.25. There is a notable difference in the magnitude of the dependence parameters in both states. In low dependence state the copula parameter is 0.5293, and in the high dependence state the parameter is 0.8184. In Figure 2.10 of Chapter 2 we observed that the high dependence state often coincided with financial crises, hence we can label this state as a financial crisis state. In addition, states seem to be persistent as suggested by high parameter estimate for the probability of transiting from *state 1* to *state 1*, p_{11}, and low parameter estimate for the probability of transiting from *state 2* to *state 1*, p_{21}.

Typically, one would be interested in carrying out inference on the estimated copula and the transition probability matrix parameters. In this situation, one would need the critical value for the test statistic, which requires the knowledge of the distribution of the test statistics under the null hypothesis. However, the true distribution of the statistic is often unknown, or may be analytically intractable. Therefore, one would generally proceed by using approximations of the underlying distribution based on asymptotic theory. Nevertheless, approximations assume that the sample size is sufficiency large so that the test statistic converges to the relevant limiting distribution. In the context of regime-switching copula models, increasing the sample size implicitly assumes that observations from both states also increase in proportion. This essentially means that the transition probability matrix P remains fixed.

Previous sections considered two alternative asymptomatic sequences that relax the assumption of a constant transition probability matrix. In this section we are interested in applying the alternative asymptotic sequence Case 1. That is, a situation in which the expected number of time periods spent in one of the states is constant as the sample size increases. This scenario is fairly conceivable in the context of financial crises. If the

government regulatory policy becomes more effective in preventing future financial crises, one would expect them to be rare over time.

Similar to previous sections, we construct a framework to incorporate these types of scenarios. Using the parameter estimates from Table 3.25, and rounding to two decimal digits, the transition probability matrix P is of the following form:

$$P = \begin{pmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{pmatrix}$$

Using equation (3.39) for a given sample size under consideration $T_0 = 1,229$, the resulting G matrix is:

$$G = \begin{pmatrix} 25.98 & -25.98 \\ 1273.02 & -1273.02 \end{pmatrix}$$

However, in order to implement the framework under the alternative asymptotic sequence Case 1, some further modification to G is required. This is because the elements of the first column of the P matrix are different, and hence the resulting diagonal elements of matrix G are also different. This means that equation (3.41) does not simplify to $\pi_2^T = \frac{g_{11}}{T+g_{11}+g_{22}} = \frac{g_{11}}{T}$. If $g_{11} = -g_{22}$, then multiplying π_2^T by T would lead to a constant expected occupancy time in crisis state, i.e. equal to g_{11}. If $g_{11} \neq -g_{22}$, then the expected occupancy time will decrease as the sample size increases, and will approach g_{11} in the limit. In order to keep the expected occupancy time constant, the following modification is applied to the G matrix:

$$G^T = \begin{pmatrix} g_{11} & g_{12} \\ g_{21}^T & g_{22}^T \end{pmatrix}$$

where $g_{22}^T = \left(\frac{g_{11}}{\pi_2^0 T_0} - 1\right) T - g_{11}$, π_2^0 is the ergodic (or steady state) probability of being in *state 2* (i.e. financial crisis), $T_0 = 1,299$ is the original sample size, T is an arbitrary sample size we are interested in such that $T_0 \leq T$, and $g_{21}^T = -g_{22}^T$. This specification ensures that, given initial sample size T_0, the expected occupancy time in *state 2* remains constant as the sample size T increases to infinity. Using equation (3.39) and the modified G^T matrix,

Table 3.25: Parameter estimates of the regime-switching copula.

θ_1	θ_2	p_{11}	p_{21}
0.5293	0.8184	0.9814	0.0186
(0.0277)	(0.0147)	(0.0079)	(0.0084)

Note: In brackets are the standard errors.

the resulting P matrix for any T is of the following form:

$$P^T = \begin{pmatrix} (1 - T^{-1}g_{11}) & -T^{-1}g_{12} \\ \left(\frac{1}{\pi_2^0 T_0} - \frac{1}{T}\right) g_{11} & \left(\frac{1}{T} - \frac{1}{\pi_2^0 T_0}\right) g_{11} + 1 \end{pmatrix}$$

where the following restrictions have to be placed on elements of the G^T matrix to ensure that the transition probabilities are non-negative and lie between 0 and 1:[10]

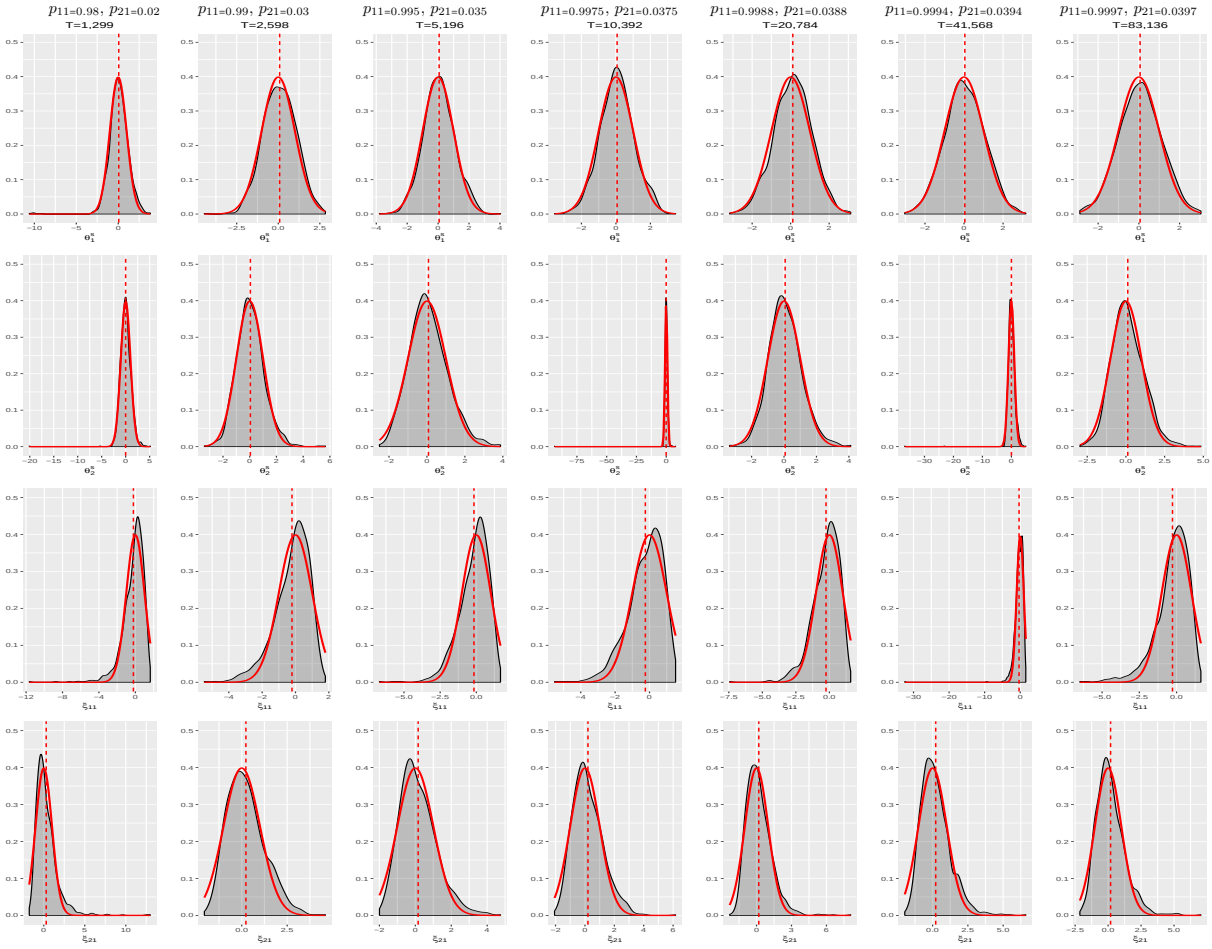
1. $g_{11} > 0$, $g_{21}^T > 0$, $g_{12} < 0$ and $g_{22}^T < 0$.
2. $g_{11} + g_{12} = g_{21}^T + g_{22}^T = 0$.

Next, we demonstrate that for parameters of the crisis state, under the modified alternative asymptotic sequence Case 1, the standard asymptotic theory may provide a poor approximation of the distribution of the test statistic under the null hypothesis even as $T \to \infty$. We consider the classical t-test. In sufficiently large samples, a confidence level α test would be to reject $H_0 : \theta = \theta_0$ if $\left|\frac{(\hat{\theta}-\theta)}{se(\hat{\theta})}\right|$ exceeds the upper $\alpha/2$ quantile of the $N(0,1)$ distribution.

First, we simulate data of length T with $N = 1,000$ replications from the regime-switching copula model in (3.37) using the parameter estimates from Table 3.25. In essence, we assume that after the time period under consideration, financial crises will not occur, which means that the expected number of time periods spent in crisis state will be stochastically bounded. Next, we calculate the test statistic from the simulated data for all parameters of interest. Figure 3.27 displays the sampling distributions of the test statistics for all parameters when the null hypothesis is true. The distribution of the t-statistic for copula parameter in non-crisis regime is displayed in the first row. As we

[10]Similar to previous cases, there still exist certain combinations of run length T and matrix G such that the resulting P_T matrix is negative. There is a simple modification to matrix P_T ensuring non-negativity, the details of which can be found in Appendix A.2.1.

Figure 3.27: Sampling distribution of the t-statistic θ_1^s, θ_2^s, ξ_{11}, and ξ_{21}.



Note: The superimposed red line is the standard Normal distribution. $\theta_1^s = (\hat{\theta}_1 - \theta_1)/\hat{se}(\hat{\theta}_1)$, $\theta_2^s = (\hat{\theta}_2 - \theta_2)/\hat{se}(\hat{\theta}_2)$ $\xi_{11} = (\hat{p}_{11} - p_{11})/\hat{se}(\hat{p}_{11})$, and $\xi_{21} = (\hat{p}_{21} - p_{21})/\hat{se}(\hat{p}_{21})$. Number of replications is N = 1,000

move across columns, the sample size increases, and the true sampling distribution seems to be well approximated by the standard normal distribution. From the last two rows of Table 3.26, we can observe that the 0.025 and 0.975 quantiles of the sampling distribution are not very far from the quantiles of the reference standard normal distribution, −1.96 and 1.96 respectively.

These results also seem to indicate negligible parameter estimate bias of θ_1. In the second row of Figure 3.27, the sampling distribution of the t-statistic for copula parameter in crisis regime does not seem to be settling down as the sample size increases. It remains reasonably skewed with slight kurtosis, although the interquartile range seems to be close to reference distribution. This suggests that the centre of the sampling distribution can be well approximated by the standard normal. In addition, there seem to be some unusual draws from time to time that induce strong skewness in the sampling distribution. Not

Table 3.26: Summary statistics for the sampling distribution of the t-statistic θ_1^s, θ_2^s, ξ_{11}, and ξ_{21}.

	T=1,299	**T=2,598**	**T=5,196**	**T=10,392**	**T=20,784**	**T=41,568**	**T=83,136**
	p_{11} = 0.98 p_{21} = 0.02	p_{11} = 0.99 p_{21} = 0.03	p_{11} = 0.995 p_{21} = 0.035	p_{11} = 0.9975 p_{21} = 0.0375	p_{11} = 0.9988 p_{21} = 0.0388	p_{11} = 0.9994 p_{21} = 0.0394	p_{11} = 0.9997 p_{21} = 0.0397
θ_1^s							
Mean	0.0629	0.0828	0.0664	0.0698	0.1114	0.0399	0.0607
Median	0.0549	0.0813	0.0453	0.0500	0.1342	0.0189	0.0635
Interquartile	1.3338	1.3830	1.3119	1.2616	1.3097	1.3244	1.4154
Std	1.2159	0.9894	1.0102	0.9780	0.9757	1.0054	1.0335
Skewness	-2.1145	-0.0836	0.0730	0.0238	-0.0737	0.0026	0.0025
Kurtosis	20.8740	3.0489	3.2136	3.1237	3.0111	2.8507	2.8932
Min	-10.6080	-4.4439	-3.7932	-3.5566	-3.1686	-3.0350	-2.8975
Max	3.8261	2.8339	4.0357	3.4589	3.1191	3.1615	3.0465
0.025 quantile	-1.9272	-1.8860	-1.9328	-1.8345	-1.8163	-1.9017	-1.9124
0.975 quantile	2.2496	1.9834	2.1165	2.1054	1.9863	1.9720	2.0394
θ_2^s							
Mean	-0.0007	0.0493	0.0811	-0.0638	0.0699	0.0057	0.1195
Median	-0.0292	-0.0231	-0.014	-0.0099	0.0109	-0.0397	0.0347
Interquartile	1.3127	1.302	1.3155	1.2784	1.2701	1.3232	1.3471
Std	1.2285	1.0428	1.008	3.1441	1.0179	1.7371	1.0446
Skewness	-4.1336	0.4933	0.4805	-26.467	0.3123	-11.7179	0.4366
Kurtosis	74.208	4.4586	3.4015	790.2427	3.694	234.077	3.5971
Min	-20.104	-3.393	-2.503	-93.7566	-3.351	-36.8022	-2.9759
Max	5.1386	5.6613	3.8779	7.8481	4.1094	4.9649	4.8427
0.025 quantile	-1.8369	-1.7684	-1.7118	-1.9727	-1.8331	-1.8019	-1.7180
0.975 quantile	2.1015	2.3100	2.2468	2.1467	2.2473	2.4404	2.3449
ξ_{11}							
Mean	-0.2188	-0.2111	-0.1633	-0.2422	-0.2515	-0.3236	-0.274
Median	0.0377	-0.0077	0.0292	-0.0473	-0.0843	-0.1641	-0.058
Interquartile	1.3718	1.3607	1.3229	1.3701	1.3369	1.3824	1.3321
Std	1.2155	1.0518	1.0253	1.0359	1.0133	1.4848	1.1233
Skewness	-2.403	-1.0425	-1.163	-0.9563	-1.1759	-10.6674	-1.4093
Kurtosis	15.7342	4.4246	5.9031	4.2959	6.1298	222.2656	6.1765
Min	-11.65	-5.4684	-6.6947	-5.5252	-7.4367	-32.461	-6.4936
Max	1.6336	1.793	1.6704	1.5177	1.5879	1.566	1.6321
0.025 quantile	-3.1441	-2.8203	-2.4699	-2.6390	-2.6741	-2.9014	-3.1159
0.975 quantile	1.3007	1.2856	1.3077	1.2544	1.1615	1.2209	1.2373
ξ_{21}							
Mean	0.2989	0.2291	0.1583	0.2036	0.1941	0.2207	0.1885
Median	0.0124	0.115	0.0025	0.0543	0.0537	0.0596	0.0266
Interquartile	1.3697	1.3897	1.3518	1.3734	1.353	1.2639	1.3285
Std	1.3666	1.0331	1.0264	1.0317	1.0869	1.0496	1.0903
Skewness	2.9942	0.623	0.7954	0.7501	1.4179	1.1715	1.18
Kurtosis	21.0444	3.2676	3.9549	4.1134	8.0925	5.844	6.1493
Min	-1.7701	-2.057	-1.9974	-2.0673	-2.3041	-1.974	-2.141
Max	12.9879	4.603	4.7411	6.1679	8.0763	6.6052	7.0364
0.025 quantile	-1.2658	-1.4223	-1.4721	-1.4327	-1.3745	-1.3439	-1.4194
0.975 quantile	3.5035	2.4971	2.5537	2.5507	2.6243	2.6262	2.6112

Note: Number of replications is N = 1,000. $\theta_1^s = (\hat{\theta}_1 - \theta_1)/\hat{se}(\hat{\theta}_1)$, $\theta_2^s = (\hat{\theta}_2 - \theta_2)/\hat{se}(\hat{\theta}_2)$, $\xi_{11} = (\hat{p}_{11} - p_{11})/\hat{se}(\hat{p}_{11})$, and $\xi_{21} = (\hat{p}_{21} - p_{21})/\hat{se}(\hat{p}_{21})$.

surprisingly, the 0.025 and 0.975 quantiles presented in Table 3.26 seem to be considerably far from the quantiles of the reference distribution for any given sample size. Therefore, these results suggest that regardless of the sample size, the exact and asymptotic p-values will be different, which can lead to contradictory conclusions about the hypothesis of interest. The plots for t-statistics associated with transition probability parameters p_{11} and p_{21} exhibit strong negative and positive skewness, respectively. The corresponding quantiles also seem to be considerably far away from the reference distribution.

3.5 Conclusion

In this study we examined through a Monte Carlo study finite properties of the maximum likelihood estimator of a Markov regime-switching Gaussian copula processes, where the transition probability matrix is local to absorbing states. Of particular interest was to examine the finite sample properties of copula parameter estimates in both states under the alternative asymptotic sequences. Two alternative asymptotic sequences have been considered: first where the transition probability matrix approaches a matrix in which the first column consists of elements equal to one, and the second column consists of elements equal to zero; and second, where transition probability matrix converges to an identity matrix.

In Case 1 of the alternative asymptotic sequence, the simulation results revealed that when states can be observed, the standard asymptotic theory works well for the rescaled sampling error in θ_1. Nevertheless, the standard asymptotic theory does not apply for the rescaled sampling error in θ_2. On the other hand, given that we choose a diagonal for which the expected number of time periods spent in *state 2* is not below 100, the behaviour does not seem to be far from the standard normal distribution. We have also considered the impact due to the lack of observability on the behaviour of the rescaled sampling errors. This created additional complexities in the estimation procedure, which resulted in the "contamination" of the behaviour in the rescaled sampling error in θ_1 in *state 1*. In particular, for a given sample size T, the deterioration was detected in the behaviour of the rescaled sampling error in θ_1. The closer the G matrix was to zero, the stronger and the more pronounced the deterioration was. It was proposed that this was due to the result of having negligible or, in some realisations, none of the observations from *state 2* when G was very close to zero. Subsequently, the distortions in the behaviour of the rescaled sampling error in θ_2 would "contaminate" the behaviour of the rescaled sampling error in θ_1. Nonetheless, as T increases to infinity, and as long as we stay sufficiently far away from the region where G was close to zero, the regular asymptotic theory is expected to work well, although non-trivial mathematics would be required to conclude about the limiting distribution.

In the alternative asymptotic sequence Case 2, where the transition matrix approaches an identity matrix, the standard asymptotic theory works well for copula parameters in both states. As T increases to infinity, the rescaled sampling errors in θ_1 and θ_2 are well behaved and converge to the standard normal distribution. This is because the expected number of time periods spent in each regime increases to infinity as T goes to infinity and, therefore, the regular asymptotic theory works well. Furthermore, for a given run length T, the lack of observability does not seem to substantially "contaminate" the behaviour of the rescaled sampling error in θ_1, even when the matrix G is very close to zero.

Simulation results suggest that if the regime switching state of the world is described by the alternative asymptotic sequence Case 1, computation of the covariance matrix based on the standard asymptotic theory could be inaccurate. If the regime switching state of the world is described by the alternative asymptotic sequence Case 2, then the usual asymptotic theory works well, and if the sampling error is standardised correctly using the usual asymptotic formulae for the covariance matrix, then it would be asymptotically distributed as a standard normal.

We have also applied the developed framework to the real data, in order to examine the finite sample distribution of the t-statistic under the alternative asymptotic sequence Case 1. These results suggest that regardless of how large the sample size is, the exact and asymptotic p-values will be different. The implication of these findings is that relying on standard asymptotic approximations could lead to erroneous conclusions about the hypotheses of interest.

Appendix A.1

A.1.1 Standard errors of the estimates using Godambe information

Let $\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_T}$ be a random sample from a density g, and let the realizations be $\boldsymbol{y_1}, \ldots, \boldsymbol{y_T}$. Let also $\boldsymbol{\Psi}$ be a vector of functions with the same dimension as parameter vector $\boldsymbol{\alpha}' = (\boldsymbol{\delta}_1', ..., \boldsymbol{\delta}_d', \boldsymbol{\theta}')$. The vector of *inference functions* is:

$$\sum_{t=1}^{T} \boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{y}_t) \tag{A1.1}$$

Suppose $\tilde{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}}(\boldsymbol{y_1}, ..., \boldsymbol{y_T})$ is the only root satisfying:

$$\sum_{t=1}^{T} \boldsymbol{\Psi}(\tilde{\boldsymbol{\alpha}}, \boldsymbol{y}_t) = \mathbf{0} \tag{A1.2}$$

also suppose $\boldsymbol{\alpha}^*$ is the only root satisfying:

$$\mathbb{E}_g\left[\boldsymbol{\Psi}(\boldsymbol{\alpha}^*, \boldsymbol{Y})\right] = \mathbf{0} \tag{A1.3}$$

Assuming the regularity conditions of score equations in asymptotic maximum likelihood theory hold for $\boldsymbol{\Psi}$. Then we have:

$$T^{1/2}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \longrightarrow \mathbf{N}(\mathbf{0}, \mathcal{G}(\boldsymbol{\alpha}^*)^{-1}) \tag{A1.4}$$

where $\mathcal{G}(\boldsymbol{\alpha}^*)$ is the Godambe information matrix (Godambe, 1960), where the inverse is given by:

$$\mathcal{G}(\boldsymbol{\alpha}^*)^{-1} = \mathbf{H}^{-1}\mathbf{J}(\mathbf{H}^{-1})' \tag{A1.5}$$

where

$$\mathbf{H} = -\mathbb{E}_g\left[\frac{\partial \boldsymbol{\Psi}(\boldsymbol{\alpha}^*, \boldsymbol{Y})}{\partial \boldsymbol{\alpha}'}\right], \quad \mathbf{J} = \mathbb{E}_g\left[\boldsymbol{\Psi}(\boldsymbol{\alpha}^*, \boldsymbol{Y}), \boldsymbol{\Psi}'(\boldsymbol{\alpha}^*, \boldsymbol{Y})\right] \tag{A1.6}$$

Due to the recursive nature of the log-likelihood for the Markov regime-switching mod-

els with unobserved states, the analytical expressions for **H** and **J** in (A1.6) cannot be obtained. Therefore, the numerical estimates have been used instead.

Appendix A.2

A.2.1 Ensuring non-negativity of the transition probability matrix P_T under the alternative asymptotic sequence

In this section we address the possibility of obtaining negative elements of a probability matrix for particular combinations of matrix G and run length T . We consider an alternative asymptotic sequence Case 1 where the transition matrix converges to a matrix in which all elements of the first column are equal to one, and all elements of the second column are equal to zero. We can define transition matrix P as a function of g_1, g_2 and T:

$$P_T = \begin{pmatrix} e^{-\frac{g_1}{T}} & (1 - e^{-\frac{g_1}{T}}) \\ e^{-\frac{g_2}{T}} & (1 - e^{-\frac{g_2}{T}}) \end{pmatrix}$$

In this specification it is evident that, in order for the transition probabilities to be positive and lie between 0 and 1, the following restrictions have to be placed on g_1 and g_2:

1. $g_1 \geq 0, \quad g_2 \geq 0$

The ergodic distribution can also be derived as follows:

$$\pi_1^T = \frac{1 - (1 - e^{-\frac{g_2}{T}})}{[1 - (1 - e^{-\frac{g_1}{T}})] + [1 - (1 - e^{-\frac{g_2}{T}})]} = \frac{1}{1 + e^{\frac{g_2}{T}} - e^{\frac{g_2 - g_1}{T}}}$$

$$\pi_2^T = \frac{1 - (1 - e^{-\frac{g_1}{T}})}{[1 - (1 - e^{-\frac{g_1}{T}})] + [1 - (1 - e^{-\frac{g_2}{T}})]} = \frac{e^{\frac{g_1}{T}} - 1}{e^{\frac{g_1}{T}} + e^{\frac{g_1 - g_2}{T}} - 1}$$

Evidently, ergodic probabilities depend on T with the following limits:

$$\begin{aligned} \lim_{T \to \infty} \pi_1^T &= 1 \\ \lim_{T \to \infty} \pi_2^T &= 0 \end{aligned}$$

We can also derive the expected number of state transitions by chain between date 0 and

date T:

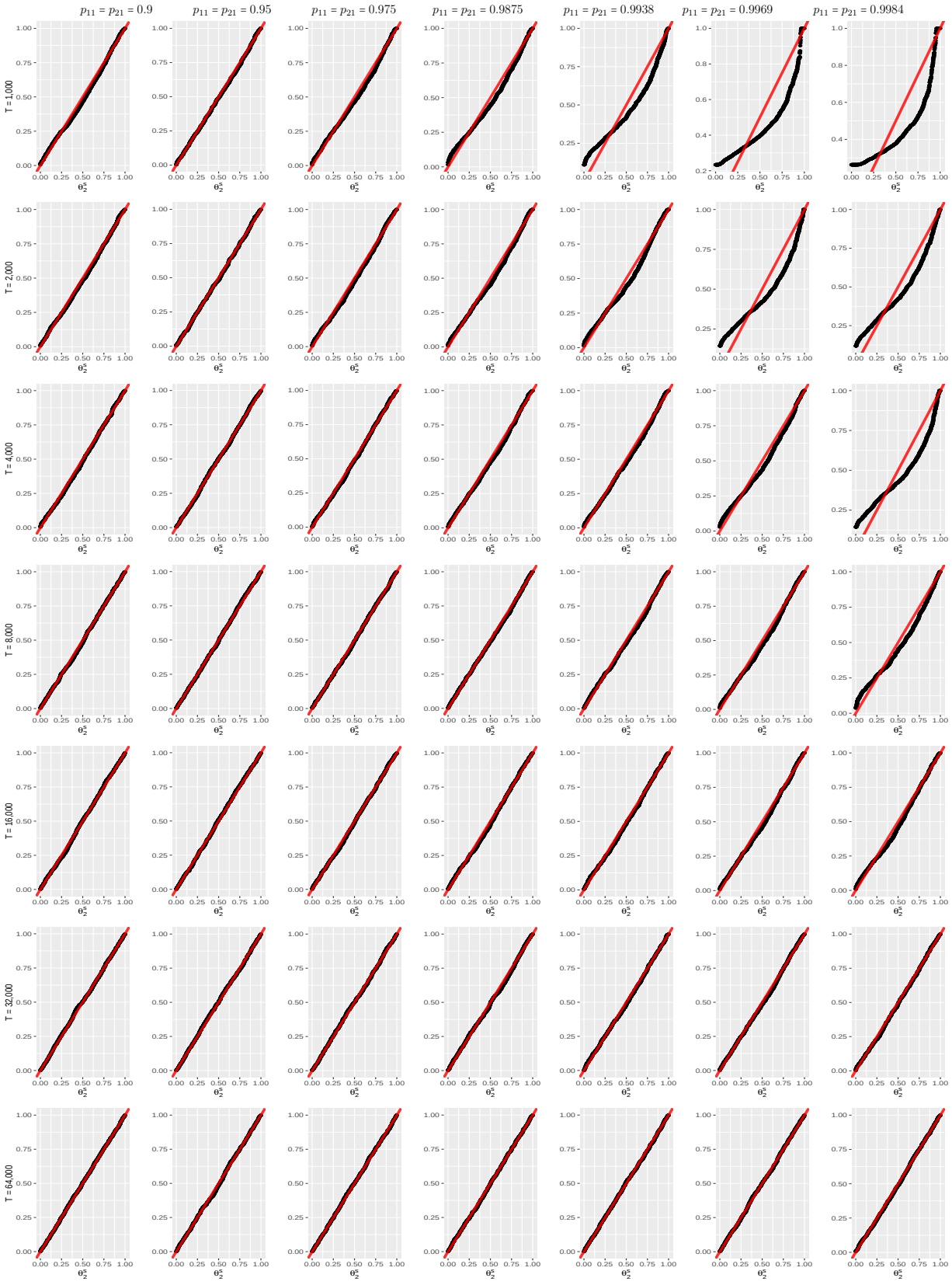
$$N_T^e = \frac{2T[(1-(1-e^{-\frac{g_2}{T}}))(1-e^{-\frac{g_1}{T}})]}{[1-(1-e^{-\frac{g_2}{T}})]+[(1-e^{-\frac{g_1}{T}})]} = \frac{2T(1-e^{-\frac{g_1}{T}})}{1+e^{\frac{g_2}{T}}-e^{-\frac{g_2-g_1}{T}}}$$

Similarly, it is evident that the expected number of state transitions also depends on T with the following limit:

$$\lim_{T\to\infty} N_T^e = 2g_1$$

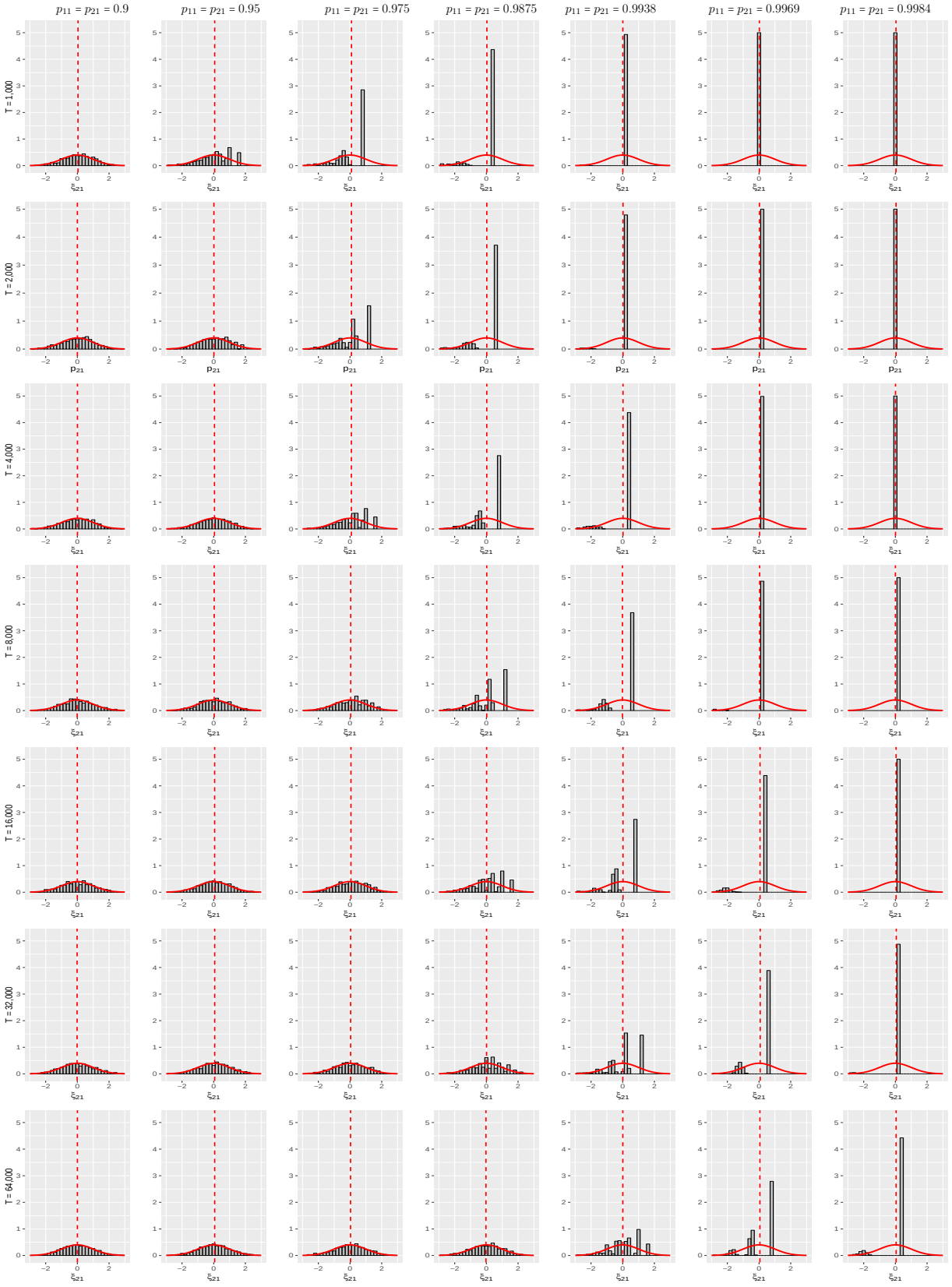
In the similar vein, the non-negativity of the transition probability matrix P_T can be ensured under the alternative asymptotic sequence Case 2, although the derivation is omitted here.

Figure 3.28: The PP-plot of the standardized distribution of estimates of θ_2^s when states are observed, (conditional on spending a positive number of time periods in State 2)



Note: $\theta_2^s = (\hat{\theta}_2 - \theta_2)/se(\hat{\theta}_2)$

Figure 3.29: Distribution of the rescaling sampling error $\xi_{21} = {}^{(\hat{p}_{21} - p_{21})}/_{se(\hat{p}_{21})}$ when states are observed, restricted X-axis. Alternative asymptotic sequence Case 1.



Note: [1] The distribution is conditional on spending a positive number of time periods in State 2. The superimposed red line is the standard Normal distribution. Number of replications is N = 1,000.

Bibliography

Aas, K. (2004): "Modelling the dependence structure of financial assets: A survey of four copulas," *Norwegian Computing Center*.

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009): "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, 44, 182–198.

Amemiya, T. (1985): *Advanced Econometrics*, Harvard University Press.

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

Ang, A. and G. Bekaert (2002): "International Asset Allocation With Regime Shifts," *Review of Financial Studies*, 15, 1137–1187.

Ang, A. and J. Chen (2002): "Asymmetric correlations of equity portfolios," *Journal of Financial Economics*, 443–494.

Bedford, T. and R. M. Cooke (2001): "Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines," *Annals of Mathematics and Artificial Intelligence*, 32, 245–268.

BIS (2010): "Basel III: A global regulatory framework for more resilient banks and banking systems," Tech. rep., Basel Committee on Banking Supervision, Bank for International Settlements (BIS).

Chan, N. H. and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Ann. Statist.*, 15, 1050–1063.

Chollete, L., A. Heinen, and A. Valdesogo (2009): "Modeling International Financial Returns with a Multivariate Regime-switching Copula," *Journal of Financial Econometrics*, 7, 437–480.

Czado, C., U. Schepsmeier, and A. Min (2012): "Maximum likelihood estimation of mixed C-vines with application to exchange rates," *Statistical Modelling*, 12, 229–255.

Doman, R. (2008): "Modeling Conditional Dependencies between Polish Financial Returns with Markov-Switching Copula Models," *Dynamic Econometric Models*, 8, 21–28.

Dufour, J.-M. and R. Luger (2017): "Identification-robust moment-based tests for Markov switching in autoregressive models," *Econometric Reviews*, 36, 713–727.

Garcia, R. and G. Tsafack (2011): "Dependence structure and extreme comovements in international equity and bond markets," *Journal of Banking Finance*, 35, 1954 – 1970.

Godambe, V. P. (1960): "An Optimum Property of Regular Maximum Likelihood Estimation," *Ann. Math. Statist.*, 31, 1208–1211.

Hamilton, J. (1994): *Time Series Analysis*, Princeton University Press.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–84.

Hayashi, F. (2011): *Econometrics*, Princeton University Press.

Joe, H. (1996): *Families of m-variate distributions with given margins and* $m(m-1)/2$ *bivariate dependence parameters*, Hayward, CA: Institute of Mathematical Statistics, vol. Volume 28 of *Lecture Notes–Monograph Series*, 120–141.

Joe, H. and J. J. Xu (1996): "The Estimation Method of Inference Functions for Margins for Multivariate Models," Tech. rep., University of British Columbia.

Jondeau, E. and M. Rockinger (2006): "The Copula-GARCH model of conditional dependencies: An international stock market application," *Journal of International Money and Finance*, 25, 827 – 853.

Kenourgios, D., A. Samitas, and N. Paltalidis (2011): "Financial crises and stock market contagion in a multivariate time-varying asymmetric framework," *Journal of International Financial Markets, Institutions and Money*, 21, 92–106.

Kim, C.-J. and C. R. Nelson (1999): *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*, Cambridge, MA: MIT Press.

Kurowicka, D. and R. Cooke (2006): *Uncertainty Analysis with High Dimensional Dependence Modelling*, John Wiley Sons, Ltd.

Long, K. (2007): "Adding Regime-Switching to a GARCH-Copula Model: Modeling the Joint Distribution of Excess Returns of S&P500 and Nasdaq Indices," *Working Paper*.

Longin, F. and B. Solnik (2001): "Extreme Correlation of International Equity Markets," .

Morales Napoles, O., R. Cooke, and D. Kurowicka (2010): "About the number of vines and regular vines on n nodes," Tech. rep., Delft University of Technology.

Patton, A. J. (2006): "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47, 527–556.

Rodriguez, J. C. (2007): "Measuring financial contagion: A Copula approach," *Journal of Empirical Finance*, 14, 401–423.

Silva Filho, O. C. d., F. A. Ziegelmann, and M. J. Dueker (2012): "Modeling dependence dynamics through copulas with regime switching," *Insurance: Mathematics and Economics*, 50, 346–356.

Sklar, A. (1959): "Fonctions de Repartition a n Dimensions et Leurs Marges," *De, Publications and l'Institut de Statistique de Paris*, 8, 229–231.

Stöber, J. and C. Czado (2014): "Regime Switches in the Dependence Structure of Multidimensional Financial Data," *Comput. Stat. Data Anal.*, 76, 672–686.