Estimation of Copula Models for Time Series, with an Application to a Model of Euro Exchange Rates

Andrew Patton

1 November, 2001.

Motivation – I

- Application of copula theory to economic problems is a fast-growing field: Rosenberg (1999) and (2000), Bouye, *et al*. (2000), Li (2000), Scaillet (2000), Embrechts, *et al*. (2001), Patton (2001a,b), Rockinger and Jondeau (2001).
- Time series dependence means that the estimation methods available in the statistics literature cannot be used
- There is a need for results on estimation of copula models for time series

Motivation – II

- The case that one variable has more data available than the other arises in many interesting cases:

1. Studies involving developed and emerging markets
2. Return on market and return on newly floated company
3. Return on market and return on company that went bankrupt
4. Studies involving euro and non-euro denominated assets

Contributions of this paper

This paper makes three main contributions:

1. We show how two-stage maximum likelihood theory may be applied to copula models for time series, extending existing statistics literature on estimation of copula models

2. We consider the possibility that the variables of interest have differing amounts of data available, and use copulas to extend existing literature

3. We investigate the small sample properties of the estimator in a simulation study.

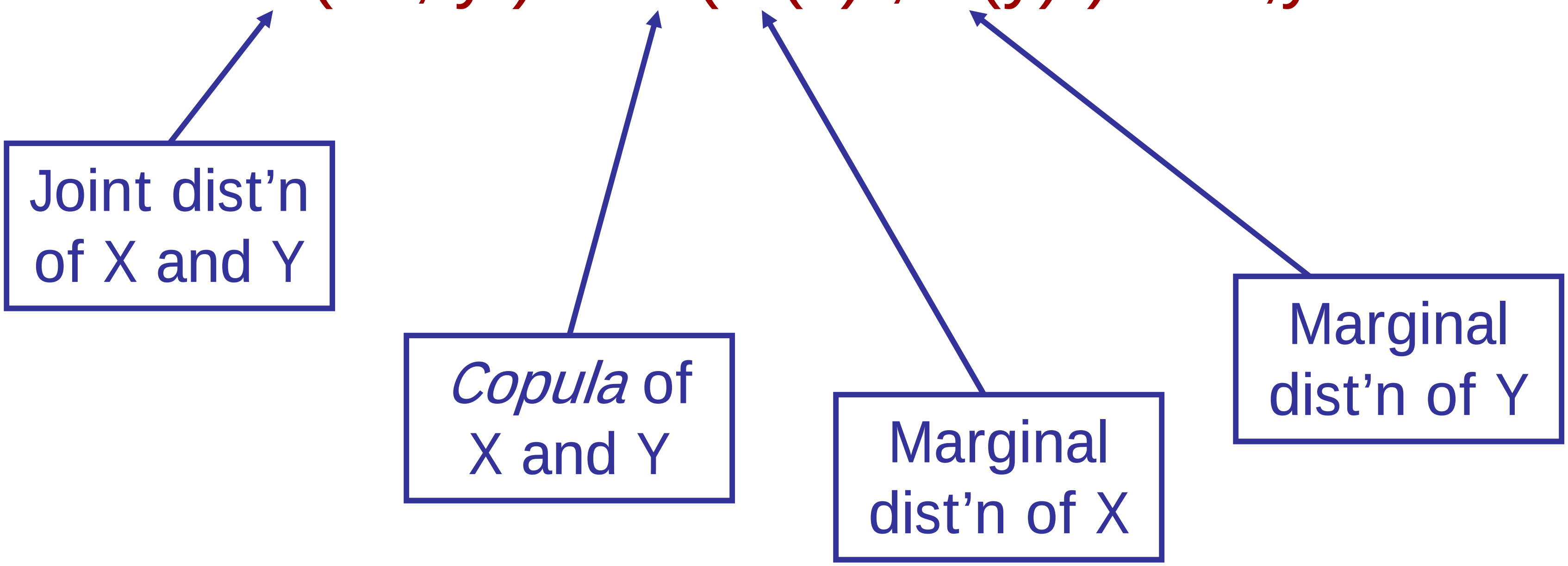
Overview

1. Refresher on copulas
2. The estimator
    i. Consistency and asymptotic normality
    ii. Covariance matrix estimation
    iii. Efficiency of the estimator
    iv. A fully efficient two-stage estimator
3. Small sample properties of the estimator
4. Application to a model of euro and yen exchange rates
5. Summary and directions of future work

Refresher on copulas

- Sklar (1959) showed that we may decompose the distribution of (X,Y) into three parts:

$$H(x, y) \Leftrightarrow C(F(x), G(y)) \quad \forall\, x,y \in \bar{\Re}$$



Refresher on copulas

- Three ways to write Sklar's theorem:

CDF:

1. $H(x, y) = C(F(x), G(y))$

PDF:

2. $h(x, y) = f(x) \cdot g(y) \cdot c(F(x), G(y))$

Log-likelihood:

3. $\log h(x, y) = \log f(x) + \log g(y) + \log c(F(x), G(y))$

so $LL_H = LL_F + LL_G + LL_C$

Log-likelihood expression

- $h(x, y) = f(x) \cdot g(y) \cdot c(F(x), G(y))$
- $LL_H = LL_F + LL_G + LL_C$

Now thinking about parametric models – consider the situation where:

- $h(x, y; \theta) = f(x; \varphi) \cdot g(y; \gamma) \cdot c(F(x; \varphi), G(y; \gamma); \kappa)$
- $LL_H(\theta) = LL_F(\varphi) + LL_G(\gamma) + LL_C(\varphi, \gamma, \kappa)$
- where $\theta = [\varphi', \gamma', \kappa']'$.

Two-stage Maximum Likelihood

$$LL_H(\theta) = LL_F(\varphi) + LL_G(\gamma) + LL_C(\varphi , \gamma , \kappa)$$

- We may exploit the fact that the parameter φ is identified in LL_F and that γ is identified in LL_G to estimate these first, and then estimate κ in LL_C *conditioning on* the estimates for φ and γ...

$\Longrightarrow$ Two-stage maximum likelihood estimation of copula models.

Relation to Anderson(1957) and Stambaugh(1997)

- Anderson (1957) and Stambaugh (1997) use the marginal/conditional decomposition:
- $h(x, y) = f(x) \cdot h_{y|x}(y|x)$, so
- $LL_H(\theta) = LL_F(\varphi) + LL_{Y|X}(\varphi, \gamma, \kappa)$
- They propose estimating φ first, and then estimating $[\gamma, \kappa]$ conditioning on the estimate of φ .
- Via the use of copulas, we are thus able to simplify estimation one further step, by breaking $LL_{Y|X}$ into the marginal likelihood of Y and the copula likelihood.

Why two-stage estimation?

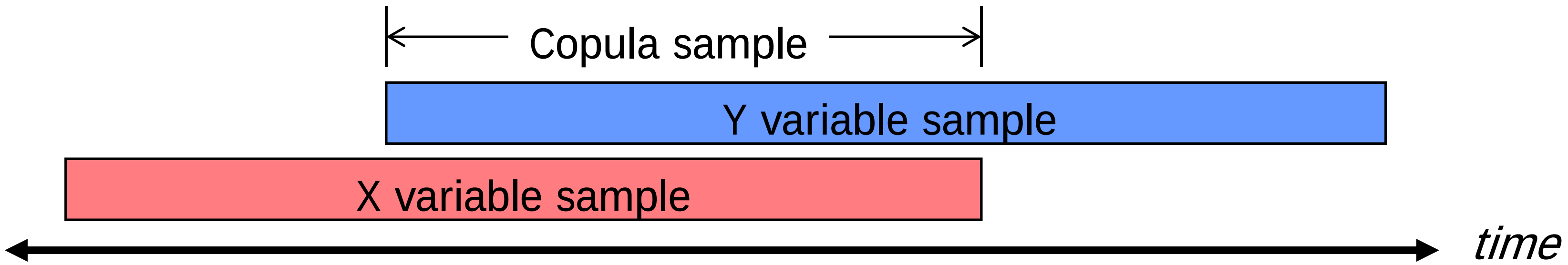
- We know (Le Cam, 1970, *inter alia*) that the (one-stage) MLE is the most efficient asymptotically normal estimator. So why think about alternatives?

1. Computational burden: for complicated models estimation becomes extremely difficult. Extension to models of higher dimension basically *requires* easier estimation methods

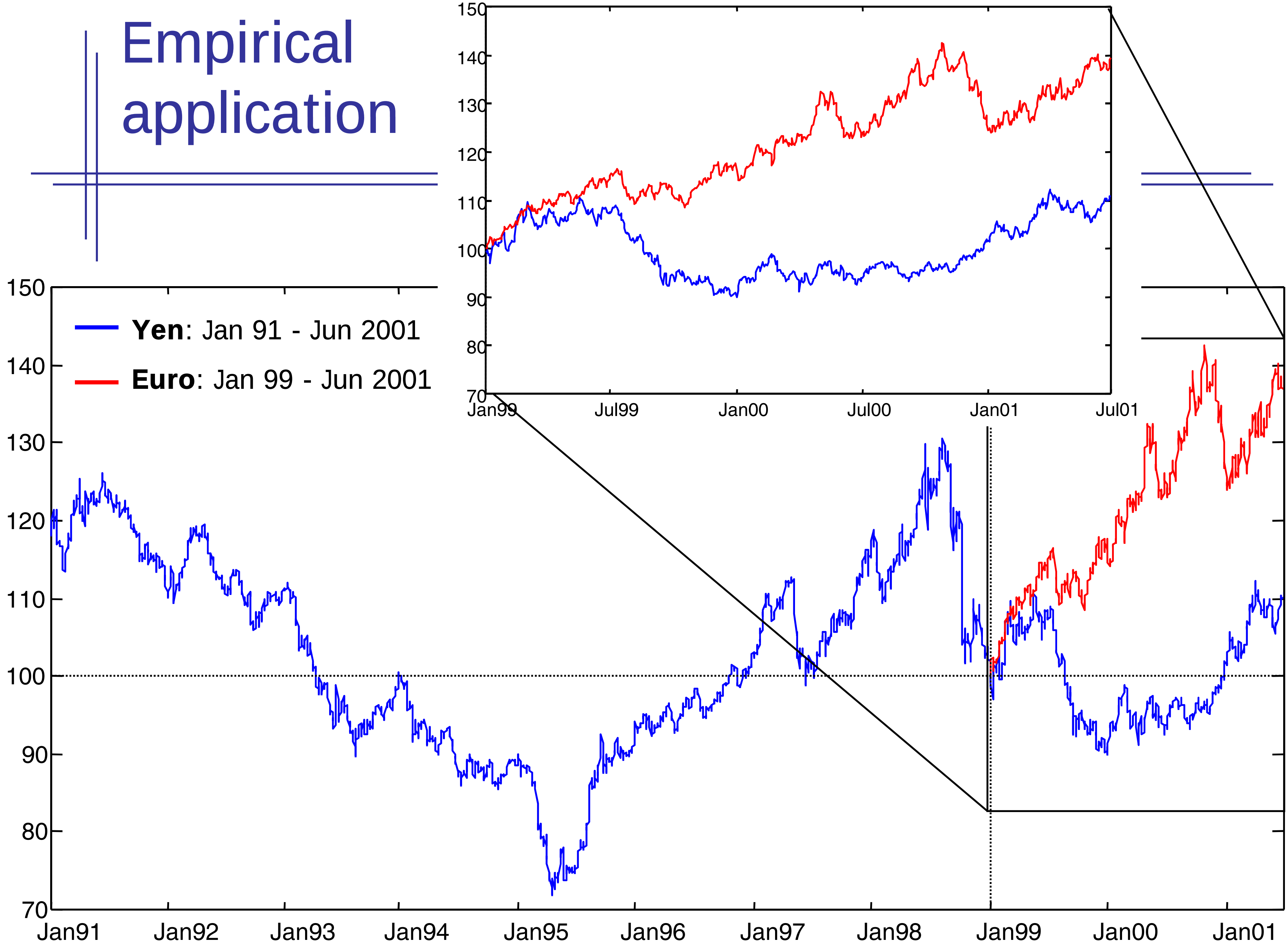
2. Modelling strategy: can work first on getting margins right, and then on copula, without iterating back and forth

Why two-stage estimation?

3. Allows for the consideration of problems with unequal amounts of data



Empirical application



Unequal data lengths

- Let the amount of data on X, Y and the copula be denoted n_x , n_y and n_c . (Note that $n_c \leq \min[n_x , n_y]$)
- We will let all of these be functions of n, and let $n_x = n$. Consider cases where $n_y / n_x \rightarrow \lambda_y \in (0,1]$ and $n_c / n_x \rightarrow \lambda_c \in (0,1]$ as $n \rightarrow \infty$
  - If $n_x - n_y$ and $n_x - n_c$ are constant as $n \rightarrow \infty$, then $\lambda_y = \lambda_c = 1$
  - If n_y / n_x and n_c / n_x are constant as $n \rightarrow \infty$, then $\lambda_y , \lambda_c \leq 1$

Two-stage maximum likelihood

$$\hat{\varphi}_{n_x} = \arg\max_{\varphi \in \Phi \subseteq R^q} n_x^{-1} \sum_{t=1}^{n_x} \log f_t(Z^t; \varphi)$$

$$\hat{\gamma}_{n_y} = \arg\max_{\gamma \in \Gamma \subseteq R^r} n_y^{-1} \sum_{t=1}^{n_y} \log g_t(Z^t; \gamma)$$

$$\hat{\kappa}_{n_c} = \arg\max_{\kappa \in \mathrm{K} \subseteq R^s} n_c^{-1} \sum_{t=1}^{n_c} \log c_t(F_t(Z^t; \hat{\varphi}_{n_x}), G_t(Z^t; \hat{\gamma}_{n_y}); \kappa)$$

$$\hat{\theta}_n \equiv [\hat{\varphi}_{n_x}, \hat{\gamma}_{n_y}, \hat{\kappa}_{n_c}]$$

Under standard conditions we obtain consistency and asymptotic normality:

Consistency result

- The use of data sets of differing lengths causes little complication for the consistency results of Newey and McFadden (1994) and White (1994) and we obtain:

$$\hat{\varphi}_{n_x} \xrightarrow{p} \varphi_0 \text{ as } n \to \infty$$

$$\hat{\gamma}_{n_y} \xrightarrow{p} \gamma_0 \text{ as } n \to \infty$$

$$\hat{\kappa}_{n_c} \xrightarrow{p} \kappa_0 \text{ as } n \to \infty$$

Asymptotic normality result

- A slight modification of the usual proof of the asymptotic normality of the two-stage MLE is required to deal with $n_x \neq n_y \neq n_c$.

$$B_n^{0-1/2} \cdot N^{1/2} \cdot A_n^0 \cdot (\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, I_s)$$

$$N \equiv \begin{bmatrix} n_x \cdot I_p & 0 & 0 \\ 0 & n_y \cdot I_q & 0 \\ 0 & 0 & n_c \cdot I_r \end{bmatrix}$$

- The asymptotic covariance matrix can be estimated using standard methods, appropriately modified.

The two-stage Hessian matrix

$$A_n^0 \equiv \begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} E[\nabla_{\varphi\varphi} \log f_t^0] & 0 & 0 \\ 0 & n_y^{-1}\sum_{t=1}^{n_y} E[\nabla_{\gamma\gamma} \log g_t^0] & 0 \\ n_c^{-1}\sum_{t=1}^{n_c} E[\nabla_{\varphi\kappa} \log c_t^0] & n_c^{-1}\sum_{t=1}^{n_c} E[\nabla_{\gamma\kappa} \log c_t^0] & n_c^{-1}\sum_{t=1}^{n_c} E[\nabla_{\kappa\kappa} \log c_t^0] \end{bmatrix}$$

The two-stage outer-product of score matrix

$$B_n^0 \equiv \mathrm{var}\left(\sum_{t=1}^{n}\left[n_x^{-1/2}s_{1t}^{0'}, n_y^{-1/2}s_{2t}^{0'}, n_c^{-1/2}s_{3t}^{0'}\right]'\right)$$

$$= \begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x}E[s_{1t}^{0}\cdot s_{1t}^{0'}] & (n_x n_y)^{-1/2}\sum_{t=1}^{n_y}E[s_{1t}^{0}\cdot s_{2t}^{0'}] & (n_x n_c)^{-1/2}\sum_{t=1}^{n_c}E[s_{1t}^{0}\cdot s_{3t}^{0'}] \\ (n_x n_y)^{-1/2}\sum_{t=1}^{n_y}E[s_{2t}^{0}\cdot s_{1t}^{0'}] & n_y^{-1}\sum_{t=1}^{n_y}E[s_{2t}^{0}\cdot s_{2t}^{0'}] & (n_y n_c)^{-1/2}\sum_{t=1}^{n_c}E[s_{2t}^{0}\cdot s_{3t}^{0'}] \\ (n_x n_c)^{-1/2}\sum_{t=1}^{n_c}E[s_{3t}^{0}\cdot s_{1t}^{0'}] & (n_y n_c)^{-1/2}\sum_{t=1}^{n_c}E[s_{3t}^{0}\cdot s_{2t}^{0'}] & n_c^{-1}\sum_{t=1}^{n_c}E[s_{3t}^{0}\cdot s_{3t}^{0'}] \end{bmatrix}$$

Asymptotic efficiency of the estimator

- When $n_x = n_y = n_c$, we know that the one-stage MLE is asymptotically most efficient

- When $n_x \neq n_y \neq n_c$ but $n_y/n_x \rightarrow 1$ and $n_c/n_x \rightarrow 1$, then $N_\infty = n \cdot I$, and one-stage is also more efficient than two-stage on data of different lengths

- But when $n_x \neq n_y \neq n_c$ and $n_y/n_x \rightarrow c < 1$ and/or $n_c/n_x \rightarrow d < 1$, then there exist cases when the two-stage estimator is *not less efficient* than the one-stage MLE...

Proposition

- Let M be the asymptotic covariance matrix of the one-stage MLE
- Let two-stage cov matrix be $V \equiv A^{-1} \cdot N^{*}_{\infty}{}^{-1/2} \cdot B \cdot N^{*}_{\infty}{}^{-1/2} \cdot A^{-1\prime}$
- Let $C = A^{-1} \cdot B \cdot A^{-1\prime}$
- Let M_{ij} denote the $(i,j)^{th}$ element of the matrix M

Prop'n: If lim $n_x / n_c \equiv d > C_{11}/M_{11}$, then the two-stage estimator obtained using all available data is *not less efficient* than the one-stage MLE.

Proposition (cont'd)

Proof: Efficiency is determined by looking at the definiteness of the asymp. covariance matrices: V-M

Proposition (cont'd)

Proof: Efficiency is determined by looking at the definiteness of the asymp. covariance matrices: V-M

Let $\lambda=[\lambda,\mathbf{0}]$, where $\lambda\in\Re/\{0\}$, then

$$\lambda'(V-C)\lambda = \lambda'(A^{-1}\cdot N^{*}_{\infty}{}^{-1/2}\cdot B\cdot N^{*}_{\infty}{}^{-1/2}\cdot A^{-1\prime} - M)\lambda$$
$$= \lambda(d^{-1}C_{11} - M_{11})\lambda$$
$$< \lambda(M_{11}/C_{11}\cdot C_{11} - M_{11})\lambda = 0$$

Proposition (cont'd)

Proof: Efficiency is determined by looking at the definiteness of the asymp. covariance matrices: V-M

Let $\lambda=[\lambda,\mathbf{0}]$, where $\lambda\in\Re/\{0\}$, then

$$\lambda'(V-C)\lambda = \lambda'(A^{-1}\cdot N^{*}_{\infty}{}^{-1/2}\cdot B\cdot N^{*}_{\infty}{}^{-1/2}\cdot A^{-1\prime} -M)\lambda$$
$$= \lambda(d^{-1}C_{11} - M_{11})\lambda$$
$$< \lambda(M_{11}/C_{11}\cdot C_{11} - M_{11})\lambda = 0$$

But, let $\lambda=[\mathbf{0},\lambda]$ where $\lambda\in\Re/\{0\}$, then

$$\lambda'(V-C)\lambda = \lambda'(A^{-1}\cdot N^{*}_{\infty}{}^{-1/2}\cdot B\cdot N^{*}_{\infty}{}^{-1/2}\cdot A^{-1\prime} -M)\lambda$$
$$= \lambda(C_{ss} - M_{ss})\lambda \quad \text{(recall C and M are sxs)}$$
$$\geq 0, \text{ by efficiency of one-stage MLE}$$

Proposition (cont'd)

Proof: Efficiency is determined by looking at the definiteness of the asymp. covariance matrices: V-M

Let $\lambda=[\lambda,\mathbf{0}]$, where $\lambda\in\Re/\{0\}$, then

$$\lambda'(V-C)\lambda = \lambda'(A^{-1}\cdot N^{*}_{\infty}{}^{-1/2}\cdot B\cdot N^{*}_{\infty}{}^{-1/2}\cdot A^{-1\prime} - M)\lambda$$
$$= \lambda(d^{-1}C_{11} - M_{11})\lambda$$
$$< \lambda(M_{11}/C_{11}\cdot C_{11} - M_{11})\lambda = 0$$

But, let $\lambda=[\mathbf{0},\lambda]$ where $\lambda\in\Re/\{0\}$, then

$$\lambda'(V-C)\lambda = \lambda'(A^{-1}\cdot N^{*}_{\infty}{}^{-1/2}\cdot B\cdot N^{*}_{\infty}{}^{-1/2}\cdot A^{-1\prime} - M)\lambda$$
$$= \lambda(C_{ss} - M_{ss})\lambda \quad \text{(recall C and M are sxs)}$$
$$\geq 0, \text{ by efficiency of one-stage MLE}$$

Thus (V-M) is indefinite. Neither estimator is more efficient than the other.

One-step efficient estimator

- Newey and McFadden (1994) and White (1994) show any asymptotically normal estimator can be made fully (asymptotically) efficient, as follows:

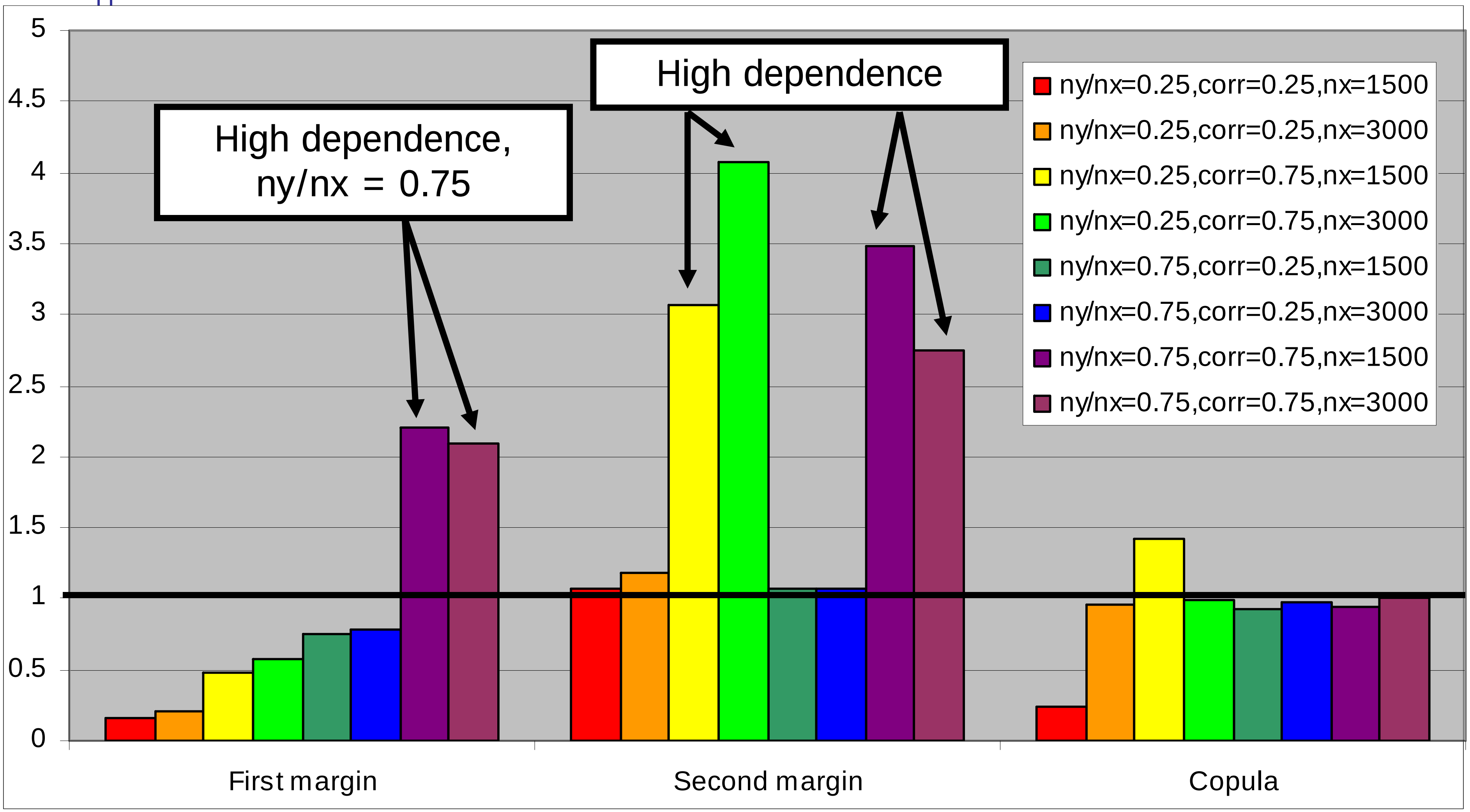
$$\hat{\theta}_n^* \equiv \hat{\theta}_n - \hat{A}_n^{-1} \cdot \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} \nabla_\varphi \log f_t(\hat{\varphi}_{n_x}) \\ n_y^{-1} \sum_{t=1}^{n_y} \nabla_\gamma \log g_t(\hat{\gamma}_{n_y}) \\ n_c^{-1} \sum_{t=1}^{n_c} \nabla_\kappa \log c_t(\hat{\theta}_n) \end{bmatrix}$$

Small sample study

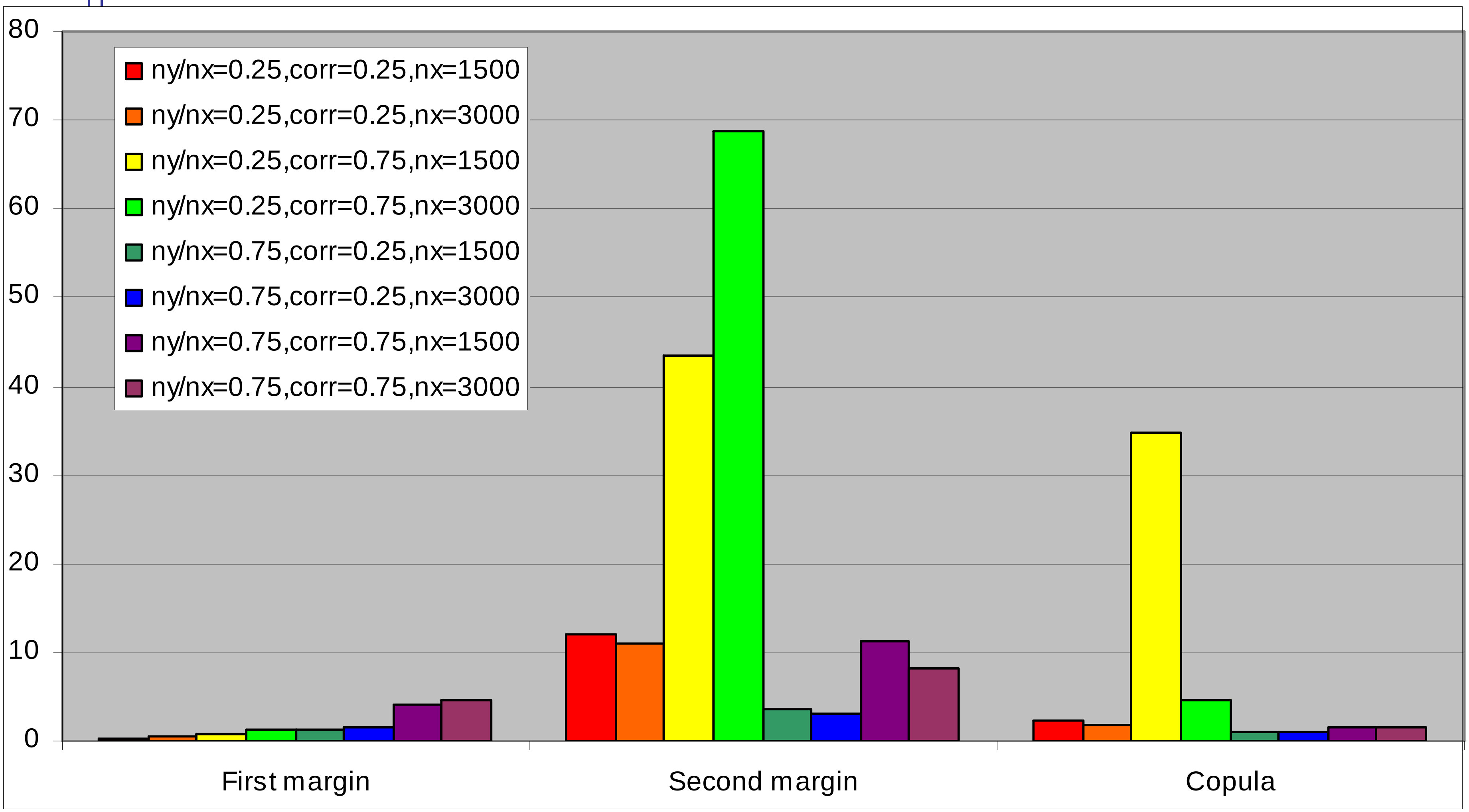
- Simulation design:

1. $(X_t , Y_t) \sim$ Clayton(Normal , Normal)
   - $X_t = 0.01 + 0.05X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,h_t^x)$
   - $h_t^x = 0.05 + 0.1\varepsilon_{t-1}^2 + 0.85h_{t-1}^x$
   - Y_t has the same specification as X_t .
2. Three dependence levels: rank correl = 0.25, 0.5, 0.75
   - Clayton copula parameters: κ = 0.41, 1.1, 2.5
3. Two lengths for n_x: n_x=1500 and n_x=3000
4. Three ratios: n_Y/n_x=0.25, 0.5 and 0.75. ($n_c=n_Y$)
5. 3 estimators: two-stage, one-step efficient, one-stage
6. 1000 replications

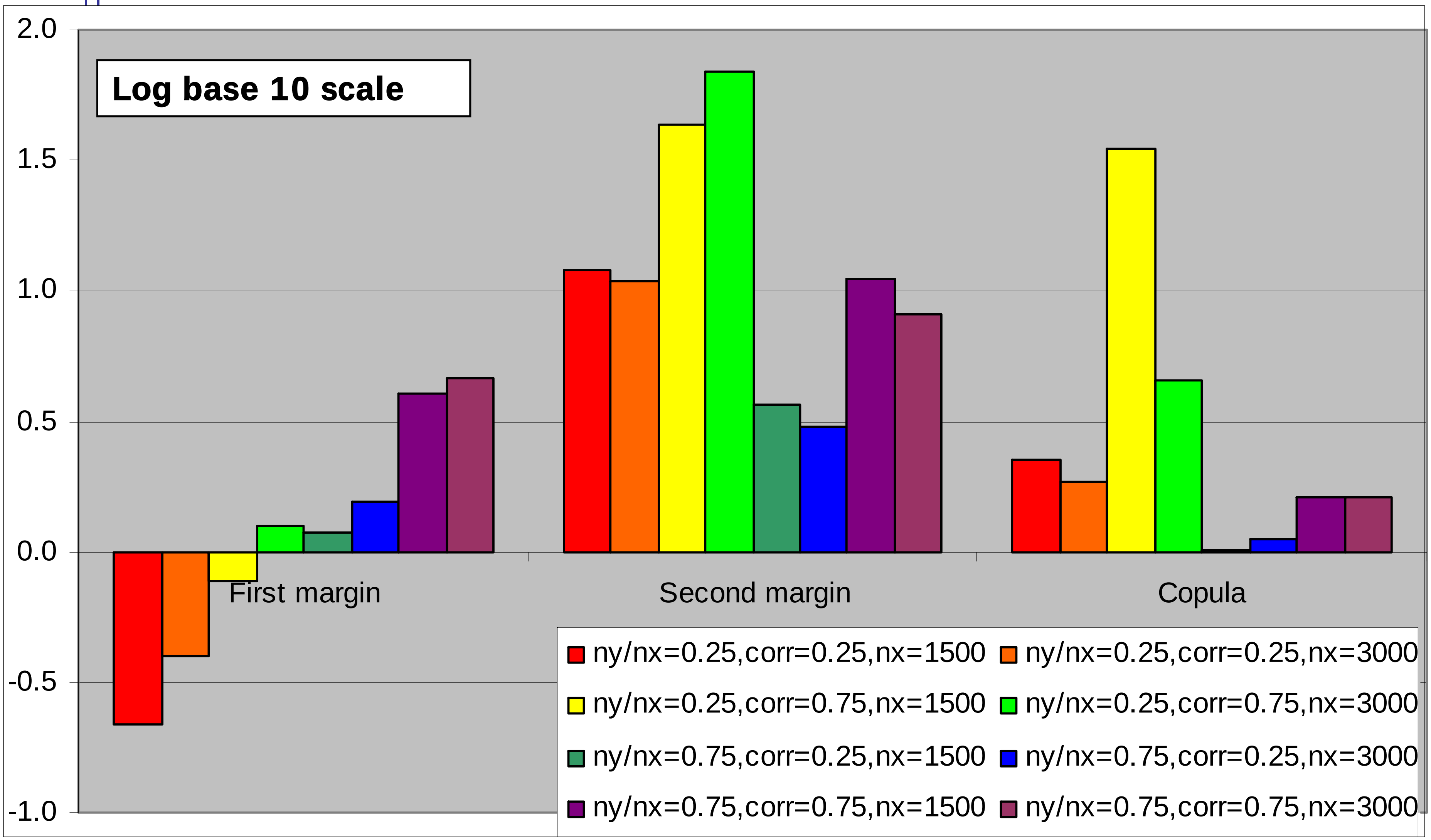
Ratio of MSEs: two-stage to one-stage



Ratio of MSEs: one-step efficient to one-stage



Ratio of MSEs: one-step efficient to one-stage

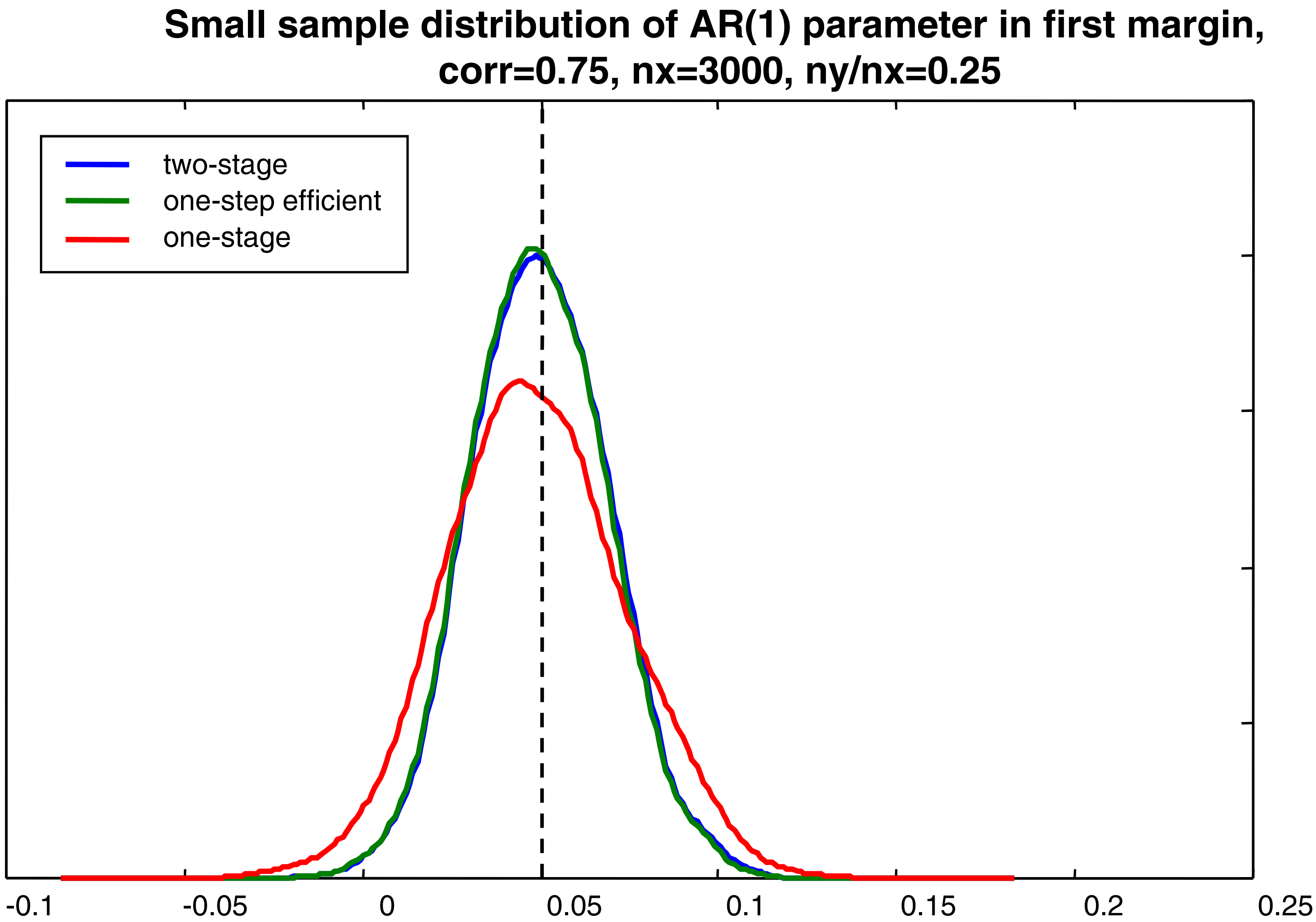


Recall: one-step efficient estimator

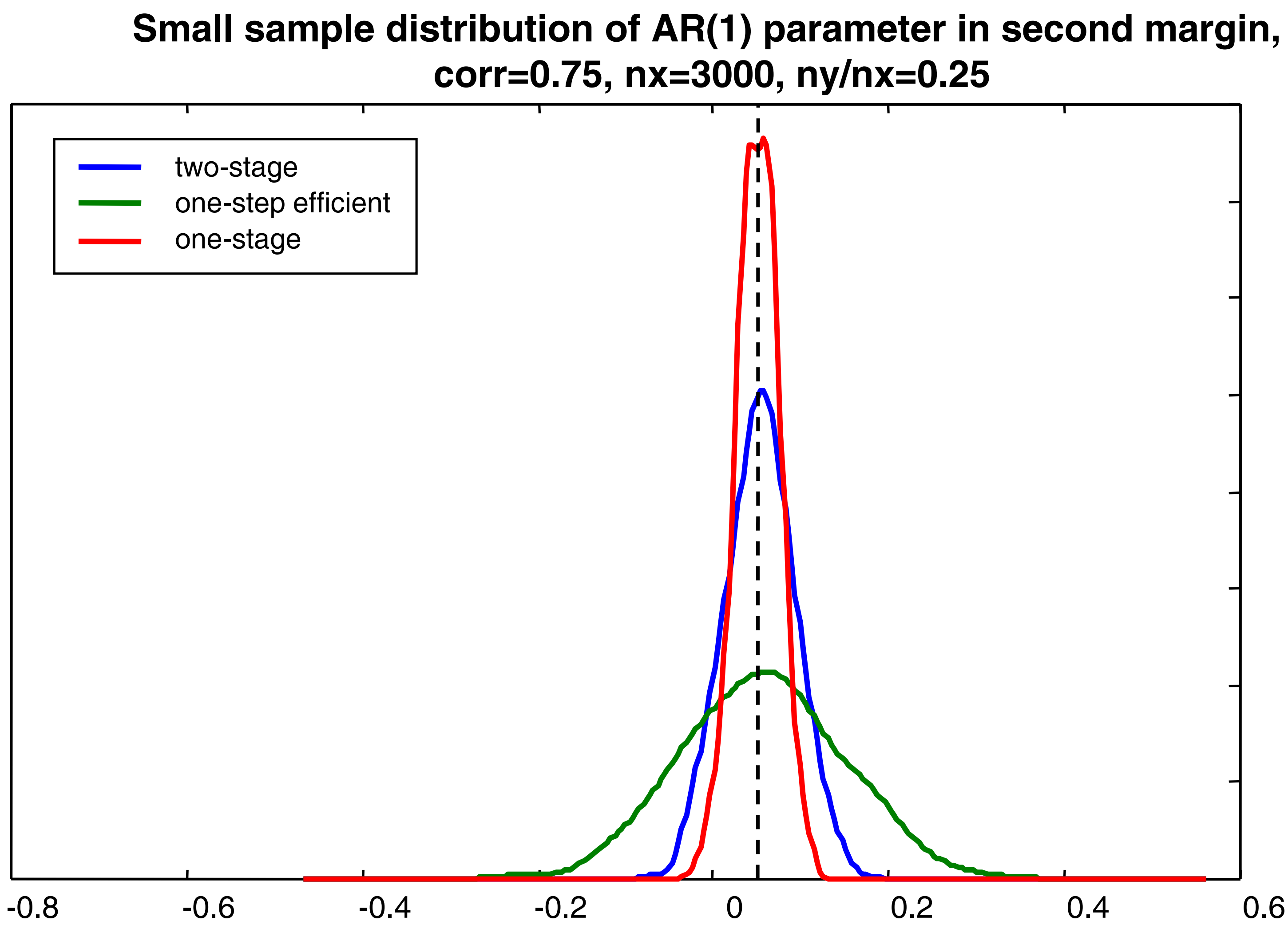
- Newey and McFadden (1994) and White (1994) show any asymptotically normal estimator can be made fully (asymptotically) efficient, as follows:

$$\hat{\theta}_n^* \equiv \hat{\theta}_n - \hat{A}_n^{-1} \cdot \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} \nabla_{\varphi} \log f_t(\hat{\varphi}_{n_x}) \\ n_y^{-1} \sum_{t=1}^{n_y} \nabla_{\gamma} \log g_t(\hat{\gamma}_{n_y}) \\ n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\kappa} \log c_t(\hat{\theta}_n) \end{bmatrix}$$

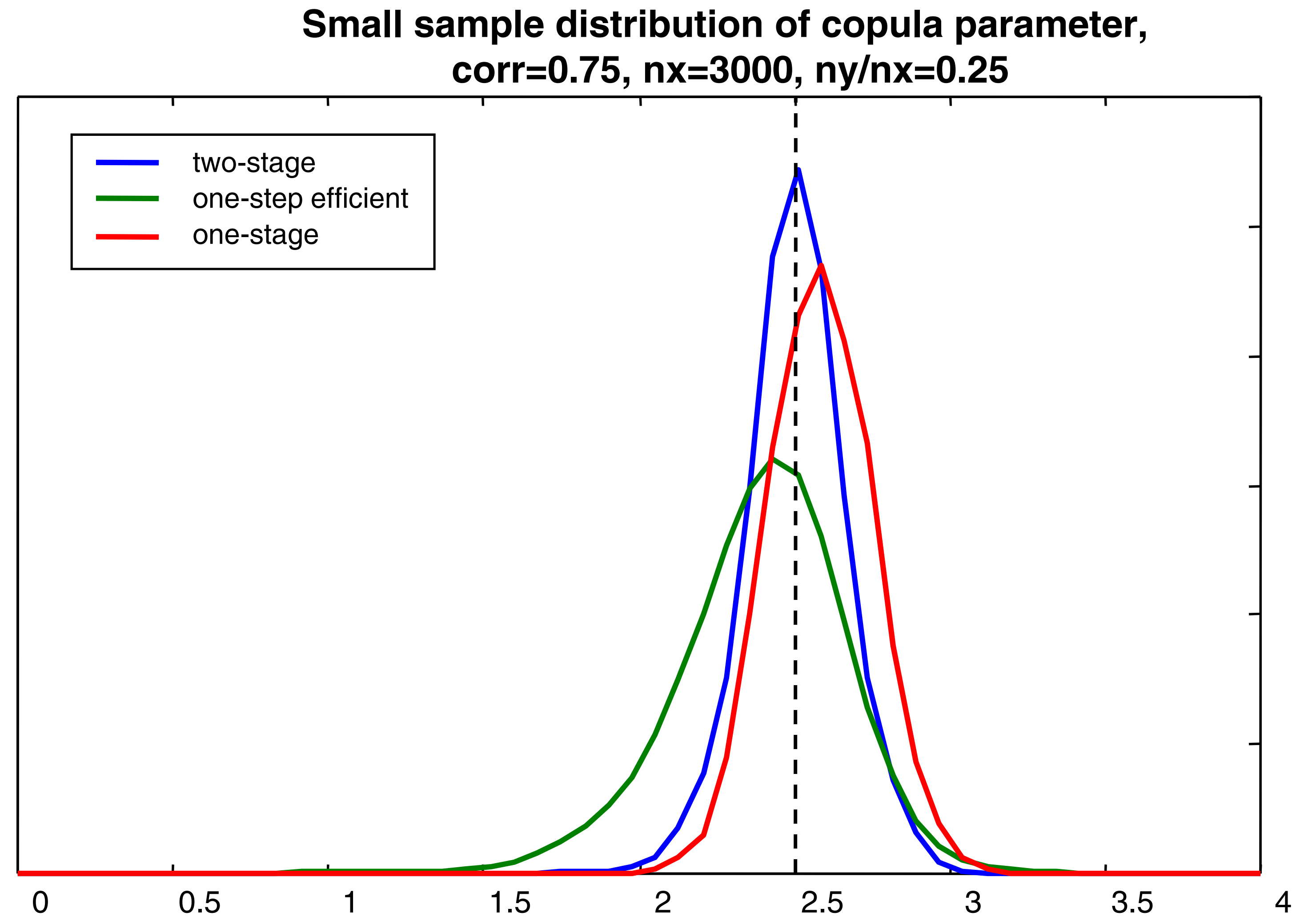
Small sample distribution of parameters



Small sample distribution of parameters



Small sample distribution of parameters



Conclusions from simulation

- Two-stage estimator performs quite well relative to the one-stage estimator
  - In many cases it has lower MSE
  - In remaining cases, the increase in MSE is moderate
- One-step efficient estimator performs quite poorly in some cases. This is attributed to the fact that it relies on an *estimate* of the covariance matrix, which amplifies small sample variability
- Overall, would recommend using unadjusted two-stage estimator rather than one-step efficient estimator

Application

- Present as an application a model of the joint distribution of yen/U.S. dollar and euro/U.S. dollar exchange rates.

- Have 2695 observations on the yen but only 643 on the euro

- Find some (weak) evidence that asymmetric Clayton copula fits better than symmetric normal and Plackett copulas.

Summary of results

- Showed how parametric copula models for time series may be estimated using two-stage maximum likelihood, easing the computational burden of MLE

- Allowed for the case of unequal amounts of data, extending existing literature.

- Presented small sample evidence that the two-stage estimator performs well relative to the one-stage estimator
  - Also found that the one-step efficient estimator had poor small sample properties in some situations