

# Improving the accuracy of bubble date estimators under time-varying volatility\*

Eiji Kurozumi<sup>a</sup>, Anton Skrobotov<sup>b,c</sup>

<sup>a</sup> Hitotsubashi University

<sup>b</sup> Russian Presidential Academy of National Economy and Public Administration

<sup>c</sup> Saint Petersburg University (Center for Econometrics and Business Analytics)

October 11, 2023

## Abstract

In this study, we consider a four-regime bubble model under the assumption of time-varying volatility and propose an algorithm of estimating the break dates with volatility correction. First, we estimate the emerging date of the explosive bubble, its collapsing date, and its recovering date to the normal market under the assumption of homoskedasticity. Second, we collect the residuals and then employ the weighted-least-squares-based estimation of the bubble dates. Using Monte Carlo simulations, we demonstrate that the accuracy of the break date estimators improves significantly via this two-step procedure in some cases compared to those based on the ordinary least squares method.

*Keywords:* rational bubble; change points; explosive autoregression; time-varying volatility; right-tailed unit root testing; mildly explosive; mildly integrated.

*JEL Codes:* C12, C22

---

\*We thank the participants at the Center for Econometrics and Business Analytics (CEBA, St. Petersburg State University) seminar series and for helpful discussions and comments. Eiji Kurozumi's research was supported by JSPS KAKENHI Grant Number 22K01422 and 23H00804. Address correspondence to Anton Skrobotov, Institute of Applied Economic Studies, Russian Presidential Academy of National Economy and Public Administration, office 2103, build 9, 82, Vernadsky pr., 117571, Moscow, Russia. E-mail: antonskrobotov@gmail.com

# 1 Introduction

Non-stationary volatility is sometimes observed in time series (in particular, financial data), but discussions on break date estimators under non-stationary volatility have received limited attention in the literature. One of the exceptions is Harris et al. (2020), wherein the estimation of level shift is improved by correcting the original time series via non-parametrically estimated time-varying variance. While the explosive bubble model has been proposed by Phillips et al. (2011) and extended by Phillips et al. (2015a,b) and Harvey et al. (2017), in which the time series is generated by a unit root process followed by an explosive regime that is again followed by a unit root regime (or with a possible stationary correction market in a recovery regime), the importance of non-stationary volatility accommodation in bubble detection methods has been discussed by Harvey et al. (2016), Harvey et al. (2018, 2020), Phillips and Shi (2020), Astill et al. (2023), and Kurozumi et al. (2023), among others. Phillips and Shi (2020) proposed a modification of the wild bootstrap recursive algorithm (based on the expanding sample) of Harvey et al. (2016) for obtaining the dates of the bubble(s) and also addressed the multiplicity testing problem. On the contrary, Harvey et al. (2020) considered the maximization of the sign-based statistic for obtaining the dates of the bubble but did not provide any finite sample performance. At the same time, as discussed in Harvey et al. (2017) and Pang et al. (2021) (PDC hereafter), the break date estimators based on the minimization of the sum of squared residuals (SSR) are more accurate than the recursive method of Phillips et al. (2015a,b) under the assumption of homoskedasticity. Nevertheless, as far as we know, there are no studies that accommodate the non-stationary volatility behavior into the estimation of the bubble dates based on the minimization of the SSR.

Recently, PDC and Kurozumi and Skrobotov (2022) investigated the asymptotic behavior of bubble date estimators. In particular, they obtained the consistency of the collapsing date estimator by minimizing the SSR using the two-regime model (even though the true model has four regimes), allowing non-stationary volatility. Owing to the consistency, one could split the whole sample at the estimated break date and consider the estimation of the date of the origination of the bubble using the sample before the estimated collapsing date and the date of the market recovery using the sample after the estimated collapsing date. This sample splitting approach closely resembles that of Harvey et al. (2017) by minimizing the full SSR based on the four-regime model, but it is computationally less involved and, as PDC demonstrated, performs better in terms of the estimation accuracy of the break dates. On the contrary to the collapsing date of the bubble, the consistency of the dates of the origination of the bubble and the market recovery depend on the extent of the explosive and

collapsing regimes. In other words, if the explosive speed is not sufficiently fast, then PDC and Kurozumi and Skrobotov (2022) obtain only the consistency of the estimators of the break fractions, not the break date.

In this study, we propose a two-step algorithm for estimating the emerging, collapsing, and recovery dates of a bubble under non-stationary volatility. First, due to the consistency of the break date (fraction) estimators regardless of heteroskedasticity, we estimate these break dates as proposed by PDC and Kurozumi and Skrobotov (2022) and collect the residuals of the fitted four-regime model. Second, we non-parametrically estimate the time-varying error variance from these residuals and perform the generalized-least-squares (GLS)-based sample splitting approach, which minimizes the weighted SSR. Monte Carlo simulations demonstrate the performance of our correction method for a model with a one-time break in volatility, especially when this break occurs at the beginning or end of the sample. The empirical application consists of different time series of cryptocurrencies for which the two methods of identifying the bubble dates are performed: one without volatility correction and another with volatility correction.

The remainder of this paper is organized as follows: Section 2 formulates the model and assumptions. In Section 3, we define the main GLS-based procedure under a general type of weights. Section 4 discusses the choice of the specific weights and proposes the new two-step algorithm. Section 5 demonstrates the finite sample performance of the estimated break dates and Section 6 provides an empirical example. Section 7 concludes the paper.

## 2 Model

Let us consider the following bubble's emerging and collapsing model for  $t = 1, 2, \dots, T$ :

$$y_t = \begin{cases} c_0 T^{-\eta_0} + y_{t-1} + \varepsilon_t & : 1 \leq t \leq k_e, \\ \phi_a y_{t-1} + \varepsilon_t & : k_e + 1 \leq t \leq k_c, \\ \phi_b y_{t-1} + \varepsilon_t & : k_c + 1 \leq t \leq k_r, \\ c_1 T^{-\eta_1} + y_{t-1} + \varepsilon_t & : k_r + 1 \leq t \leq T, \end{cases} \quad (1)$$

where  $y_0 = o_p(T^{1/2})$ ,  $c_0 \geq 0$ ,  $\eta_0 > 1/2$ ,  $\phi_a > 1$ ,  $\phi_b < 1$ ,  $c_1 \geq 0$ , and  $\eta_1 > 1/2$ . We assume that the market is normal in the first and last regimes in the sense that the time series  $y_t$  is a unit root process (a random walk) with a possibly positive drift shrinking to 0. The process starts exploding at  $t = k_e + 1$  at a rate of  $\phi_a$ , which is typically only slightly greater than one and thus sometimes characterized as a mildly explosive specification. The explosive behavior stops at  $t = k_c$  and  $y_t$  is collapsing at a rate of  $\phi_b < 1$  in the next regime, followed by the

normal market regime. This model can be regarded as a structural change model with the break points being given by  $k_e$ ,  $k_c$ , and  $k_r$ . The corresponding break fractions are defined as  $\tau_e := k_e/T$ ,  $\tau_c := k_c/T$ , and  $\tau_r := k_r/T$ , respectively. We would like to estimate these break dates as accurately as possible.

For model (1), we make the following assumptions:

**Assumption 1**  $0 < \tau_e < \tau_c < \tau_r < 1$ .

**Assumption 2**  $\varepsilon_t := \sigma_t e_t$ , where  $\{e_t\} \sim i.i.d.(0, 1)$  with  $E[e_t^4] < \infty$  and  $\sigma_t := \omega(t/T)$ , where  $\omega(\cdot)$  is a nonstochastic and strictly positive function on  $[0, 1]$ , satisfying  $\underline{\omega} < \omega(\cdot) < \bar{\omega} < \infty$ .

By Assumption 1, the break fractions are distinct and not too close each other. Assumption 2 allows for various kinds of nonstationary unconditional volatility in the shocks, such as a volatility shift (possibly multiple times) and linear and non-linear transitions. Under Assumption 2, it is well known that the functional central limit theorem (FCLT) holds for the partial sum process of  $\{\varepsilon_t\}$  normalized by  $\sqrt{T}$ , which weakly converges to a variance-transformed Brownian motion, as shown by Cavaliere and Taylor (2007a,b).

### 3 Individual Estimation of Break Dates

Following PDC and Kurozumi and Skrobotov (2022), we estimate the break dates one at a time. As model (1) can be expressed as

$$y_t = \begin{cases} \phi_1 y_{t-1} + u_t \\ \phi_a y_{t-1} + u_t \\ \phi_b y_{t-1} + u_t \\ \phi_1 y_{t-1} + u_t \end{cases} \quad \text{where } \phi_1 = 1 \quad \text{and} \quad u_t := \begin{cases} c_0/T^{\eta_0} + \varepsilon_t \\ \varepsilon_t \\ \varepsilon_t \\ c_1/T^{\eta_1} + \varepsilon_t, \end{cases} \quad (2)$$

PDC and Kurozumi and Skrobotov (2022) proposed to fit the one-time structural change model without a constant and to estimate the break point by minimizing the SSR. It is shown that the estimated break date,  $\hat{k}_c$ , is consistent for  $k_c$ . We then split the whole sample into the two subsamples, and from the first subsample before  $\hat{k}_c$ , the emerging date of the explosive behavior is estimated by fitting a one-time structural change model again, while  $k_r$  is estimated from the second subsample after  $\hat{k}_c$ . These estimated break fractions,  $\hat{\tau}_e := \hat{k}_e/T$  and  $\hat{\tau}_r := \hat{k}_r/T$ , are shown to be consistent, and furthermore,  $\hat{k}_e$  ( $\hat{k}_r$ ) is consistent for  $k_e$  ( $k_r$ ) if, roughly speaking,  $\phi_a$  deviates from 1 sufficiently ( $\phi_a - 1 > 1 - \phi_b$ ). See PDC and Kurozumi and Skrobotov (2022) for details.

Although the aforementioned estimated break dates (fractions) are consistent under non-stationary volatility in Assumption 2, the efficiency gain would be expected by estimating the break dates based on the weighted SSR following Xu and Phillips (2008). To be more precise, let  $w_t$  be a generic series of weights  $w_t$ , and then, the weighted SSR based on a one-time structural change model is given by

$$SSR(k, w_t, \phi_a, \phi_b) := \sum_1^k w_t^{-2} (y_t - \phi_a y_{t-1})^2 + \sum_{k+1}^T w_t^{-2} (y_t - \phi_b y_{t-1})^2, \quad (3)$$

where  $\sum_{t=\ell}^m$  is abbreviated just as  $\sum_\ell^m$ . As  $SSR(k, w_t, \phi_a, \phi_b)$  is minimized at

$$\hat{\phi}_a(k, w_t) := \frac{\sum_1^k y_{t-1} y_t w_t^{-2}}{\sum_1^k y_{t-1}^2 w_t^{-2}} \quad \text{and} \quad \hat{\phi}_b(k, w_t) := \frac{\sum_{k+1}^T y_{t-1} y_t w_t^{-2}}{\sum_{k+1}^T y_{t-1}^2 w_t^{-2}}$$

for given  $k$  and  $w_t$ , the estimator of  $k_c$  is given by

$$\hat{k}_c(w_t) := \arg \min_{\underline{\tau}_c \leq k/T \leq \bar{\tau}_c} SSR(k, w_t),$$

where  $0 < \underline{\tau}_c < \tau_c < \bar{\tau}_c < 1$  and  $SSR(k, w_t) := SSR(k, w_t, \hat{\phi}_a(k, w_t), \hat{\phi}_b(k, w_t))$ . The corresponding break fraction estimator is defined as  $\hat{\tau}_c(w_t) := \hat{k}_c(w_t)/T$ .

Once we obtain the estimator of  $k_c$ , we can move on to the estimation of  $k_e$  and  $k_r$ . For  $k_e$ , on the one hand, the estimation is based on the minimization of the weighted SSR using the first sub-sample, and the estimator is defined as

$$\hat{k}_e(w_t) := \arg \min_{\underline{\tau}_e \leq k/T \leq \bar{\tau}_e} SSR_1(k, w_t)$$

where  $0 < \underline{\tau}_e < \tau_e < \bar{\tau}_e < \hat{\tau}_c$  and

$$SSR_1(k, w_t) := \sum_1^k w_t^{-2} (y_t - \hat{\phi}_c(k, w_t) y_{t-1})^2 + \sum_{k+1}^{\hat{k}_c(w_t)} w_t^{-2} (y_t - \hat{\phi}_d(k, w_t) y_{t-1})^2$$

$$\text{with } \hat{\phi}_c(k, w_t) := \frac{\sum_1^k y_{t-1} y_t w_t^{-2}}{\sum_1^k y_{t-1}^2 w_t^{-2}} \quad \text{and} \quad \hat{\phi}_d(k, w_t) := \frac{\sum_{k+1}^{\hat{k}_c(w_t)} y_{t-1} y_t w_t^{-2}}{\sum_{k+1}^{\hat{k}_c(w_t)} y_{t-1}^2 w_t^{-2}}.$$

The corresponding break fraction estimator is defined as  $\hat{\tau}_e(w_t) := \hat{k}_e(w_t)/T$ . For notational convenience, we suppressed the dependence of  $\hat{k}_e(w_t)$ ,  $\hat{\tau}_e(w_t)$ , and  $SSR_1(k, w_t)$  on  $\hat{k}_c(w_t)$ .

On the other hand, for the estimation of  $k_r$ , we minimize the weighted SSR using the second sub-sample, and the estimator is defined as

$$\hat{k}_r(w_t) := \arg \min_{\underline{\tau}_r \leq k/T \leq \bar{\tau}_r} SSR_2(k, w_t)$$

where  $\hat{\tau}_c < \underline{\tau}_r < \tau_r < \bar{\tau}_r < 1$  and

$$SSR_2(k, w_t) := \sum_{\hat{k}_c(w_t)+1}^k w_t^{-2} \left( y_t - \hat{\phi}_e(k, w_t) y_{t-1} \right)^2 + \sum_{k+1}^T w_t^{-2} \left( y_t - \hat{\phi}_f(k, w_t) y_{t-1} \right)^2$$

$$\text{with } \hat{\phi}_e(k, w_t) = \frac{\sum_{\hat{k}_c(w_t)+1}^k y_{t-1} y_t w_t^{-2}}{\sum_{\hat{k}_c(w_t)+1}^k y_{t-1}^2 w_t^{-2}} \quad \text{and} \quad \hat{\phi}_f(k, w_t) = \frac{\sum_{k+1}^T y_{t-1} y_t w_t^{-2}}{\sum_{k+1}^T y_{t-1}^2 w_t^{-2}}.$$

The corresponding break fraction estimator is defined by  $\hat{\tau}_r(w_t) := \hat{k}_r(w_t)/T$ .

We call the above-stated method the sample splitting approach based on the weighted least squares (WLS) method. Note that the special case where  $w_t = 1$  for all  $t$  corresponds to the estimation method employed by PDC, with the sample ranging from 1 to  $k_r$ , and that by Kurozumi and Skrobotov (2022).

## 4 Adaptive Estimation

To implement the sample splitting approach based on the WLS method in practice, we need to choose an appropriate weight function  $w_t$ . In our model, it is natural to choose the volatility function  $\sigma_t$  as the weight  $w_t$  to obtain the efficiency gain, but such a WLS estimation is infeasible because the volatility function is unknown. In this article, we follow Xu and Phillips (2008) and estimate  $\sigma_t$  via a kernel-based method. More precisely, as the first step, we estimate  $\tau_e$ ,  $\tau_c$ , and  $\tau_r$  using the sample splitting approach based on the ordinary least squares (OLS) method ( $w_t = 1$  for all  $t$ ), as proposed by PDC and Kurozumi and Skrobotov (2022). Furthermore, using the estimated break dates denoted as  $\hat{\tau}_e(1)$ ,  $\hat{\tau}_c(1)$ , and  $\hat{\tau}_r(1)$ , we estimate

$$\Delta y_t = \mu_1 D_t(\hat{\tau}_e(1), \hat{\tau}_c(1)) + \mu_2 D_t(\hat{\tau}_c(1), \hat{\tau}_r(1)) + \phi_a D_t(\hat{\tau}_e(1), \hat{\tau}_c(1)) y_{t-1} + \phi_b D_t(\hat{\tau}_c(1), \hat{\tau}_r(1)) y_{t-1} + e_t, \quad (4)$$

via the OLS method and obtain the residuals  $\hat{e}_t$ , where  $D_t(a, b) = \mathbb{I}([\mathit{aT}] < t \leq [\mathit{bT}])$ , with  $\mathbb{I}(\cdot)$  being the indicator function. Using these first-step residuals, the second step entails the calculation of the weight function, which is used for the WLS estimator, that is, we calculate  $\hat{\sigma}_t^2$  as

$$\hat{\sigma}_t^2 = \sum_{i=1}^T \left( \sum_{i=1}^T K_{it} \right)^{-1} K_{it} \hat{e}_t^2, \quad \text{where } K_{it} = \begin{cases} K\left(\frac{t-i}{Tb}\right) & \text{if } t \neq i \\ 0 & \text{if } t = i \end{cases}, \quad (5)$$

$K(\cdot)$  is a bounded nonnegative continuous kernel function defined on the real line with  $\int_{-\infty}^{\infty} K(s) ds = 1$ , and  $b$  is a bandwidth parameter. By plugging  $\hat{\sigma}_t^2$  into  $w_t^2$  in the sample splitting approach, we obtain the estimators  $\hat{\tau}_e(\hat{\sigma}_t)$ ,  $\hat{\tau}_c(\hat{\sigma}_t)$ , and  $\hat{\tau}_r(\hat{\sigma}_t)$ . Xu and Phillips

(2008) showed that the estimation accuracy of the coefficient in a stable autoregressive model improves by the adaptive (WLS) estimation, and we investigate if their observation holds for the estimation of the bubble's dates in the next section.

## 5 Monte Carlo Simulations

In this section, we examine the performance of the estimates of the bubble regime dates in finite samples when the error variance is subject to changes in volatility.

Monte-Carlo simulations reported in this section are based on the series generated by model (1) with  $y_0 = 1500$  and  $\{\varepsilon_t\} \sim IIDN(0, 1)$ . Data are generated from this DGP for samples of  $T = (400, 800)$  with 50,000 replications.<sup>1</sup> We set the drift terms in the first and fourth regimes to  $c_0T^{-\eta_0} = 1/800$  and  $c_1T^{-\eta_0} = 1/800$ , respectively, following PDC. In this experiment, we focus on the local to unit root behavior characterized by  $\phi_a = 1 + c_a/T$  and  $\phi_b = 1 - c_b/T$ , where  $c_a$  takes values among  $\{4, 5, 6\}$  whereas  $c_b$  is fixed at 6.

For the dates of bubble regimes, we set  $(\tau_e, \tau_c, \tau_r)$  to be equal to  $(0.4, 0.6, 0.7)$ . This setting seems to be empirically relevant considering Japanese stock price, the US house price index, and cryptocurrencies. We consider the case in which there is a one-time break in volatility at date  $\tau$ , so that the volatility function  $\sigma_t$  has the following form:

$$\sigma_t^2 = s_0^2 + \delta(s_1^2 - s_0^2)\mathbb{I}(t > \lfloor \tau T \rfloor)$$

where  $s_1/s_0$  takes values among  $\{1/5, 5\}$  and  $\tau$  takes values among  $\{0.2, 0.8\}$ .

As in Kurozumi and Skrobotov (2022), in the minimization of  $SSR(k/T)$ ,  $SSR_1(k/T)$ , and  $SSR_2(k/T)$ , we exclude the first and last 5% observations from the permissible break date  $k$ . For example, when estimating  $k_r$  based on  $SSR_2(k/T)$ , the permissible break date  $k$  ranges from  $\hat{k}_c + 0.05T + 1$  to  $0.95T$ . If the break date estimate  $\hat{k}_c$  exceeds  $0.95T$ , then we cannot estimate  $k_r$ ; we do not include such a case in any bins of the histogram and thus the sum of the heights of the bins is not necessarily equal to one for  $\hat{k}_r$  in some cases. Similarly, we cannot estimate  $\hat{k}_e$  when  $\hat{k}_c < 0.05T$ . To save space, we pick up several selected cases in the following, and the other cases are provided in the online appendix.

Figure 1 presents the histograms of  $\hat{k}_c$  when  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ , and  $T = 400$ . The left-hand column shows the results based on the OLS method, while the right-hand column corresponds to the WLS-based method. In this case, the process becomes more volatile at the end of the sample, and thus, it would be difficult to distinguish between the explosive

<sup>1</sup>All simulations were programmed in R with rnorm random number generator.

and collapsing behavior (from  $\tau = 0.4$  to  $0.7$ ) and a random walk with high volatility (from  $\tau = 0.8$  to  $1$ ). As expected, the OLS method incorrectly chooses the end of the sample as the collapsing date when  $c_a = 4$ , as shown in Figure 1(a), although the local peak is observed at around the true break fraction ( $\tau_c = 0.6$ ). As the size of the bubble ( $c_a$ ) gets larger, the local peak becomes higher, as is observed in Figures 1(c) and (e) (note that the vertical axis is different depending on the value of  $c_a$ ). On the contrary, we can observe from Figures 1(b), (d), and (f) that the WLS method can estimate the collapsing date more accurately than the OLS method; the finite sample distribution has a mode at the true break fraction and the frequency of correctly estimating the true date by WLS is about twice of that by OLS.

Figure 2 shows the histograms of  $\hat{k}_c$  when  $\tau = 0.2$ ,  $s_0/s_1 = 5$ , and  $T = 400$ . In this case, there exists a unit root regime with high volatility at the beginning of the sample, and thus, it is expected that the histograms would have positive frequencies before  $\tau = 0.2$ . In fact, this is the case as observed in Figure 2, although the accuracy is much better than the case in Figure 1. Overall, the WLS-based method can detect the true collapsing date more often than the OLS-based method. For example, when  $c_a = 4$  and  $c_b = 6$ , the relative frequency of correct detection of the true collapsing date rises from  $0.25$  to  $0.35$  by introducing the adaptive procedure. We can also observe that the WLS method incorrectly detects the collapsing date at the beginning of the sample less frequently than the OLS method.

Figure 3 presents the histograms of  $\hat{k}_e$  when  $\tau = 0.2$ ,  $s_0/s_1 = 5$ , and  $T = 400$ , which is the same case as in Figure 2. Overall, when the size of the bubble is small with  $c_a = 4$ , it is difficult to estimate the emerging date ( $\tau_e = 0.4$ ) accurately, but for large values of  $c_a$ , the accuracy of  $\hat{k}_e$  improves and the histograms have a peak at  $0.4$ , as in Figures 3(c)–(f). Again, in this case, the performance of the estimator based on the WLS method is better than that based on the OLS method.

The other results are briefly summarized in the online appendix. Overall, Monte Carlo simulations demonstrate that the accuracy of the estimators of the break dates improves significantly in some cases, while in other cases, we cannot find any difference between the distribution of the estimator based on the OLS method and that based on the WLS method. Because our volatility correction does not deteriorate the finite sample performance of the break dates estimators, we recommend using the sample splitting approach with the WLS-based method in all the cases.



## 6 Empirical Application

In this section, we demonstrate the application of the two sample splitting approaches to the top largest cryptocurrencies by capitalization (btc, eth, xrp, xlm, bch, ltc, eos, bnb, ada, xtz, etc, xmr) for daily observations. In all cases, the closing price in US dollars at 00:00 GMT on the corresponding day is used. Recently, Kurozumi et al. (2023) investigated the explosive behavior of these time series and detected explosiveness as well as non-stationary volatility behavior. We implement the two estimation methods for each calendar year (365 observations from January 1 to December 31), from 2014 to 2019, if the data of the corresponding currencies are available in that year. We report only the cases where the two methods return the different estimates of the break dates, because the purpose of this section is to demonstrate the effectiveness of the WLS method in identifying the dates of the explosive behavior. Therefore, we omit the cases where the break dates are the same in both methods.

We find eight cases where at least one of the estimated break dates is different. The results are presented in Figures 4-11. In each figure, the black line shows the sample path of the corresponding cryptocurrency. The three red dotted lines indicate the estimated dates of the emergence, collapse, and recovery based on the OLS method; furthermore, the three blue dashed lines represent those estimated by the WLS method.

For xrp in 2014 in Figure 4, the series is collapsing from the beginning of the sample, and it seems to be explosive, at least through visual inspection, at the end of the sample. Clearly, our model (1) with one explosive regime is not valid in the corresponding year. In such a case, both methods cannot identify the correct break dates. This example demonstrates that we should be careful when choosing the sample periods and select those in which only one set of the four regimes should be included in the same order as in model (1).

Figure 5 shows xrm in 2015. We can observe that the same collapsing and recovering dates of the explosive behavior are obtained by the two methods, whereas the estimated emerging date by the WLS is about one month earlier than that by the OLS, especially if we take nonstationary volatility into account.

Figure 6 shows eth in 2016, which may become explosive twice by visual inspection. It seems that the WLS method successfully detects the explosive behavior of eth in early 2016, whereas the OLS method erroneously assigns the second peak of the process as the recovering date.

The currency xlm in 2017 is given in Figure 7, in which the small explosive behavior exists at the middle of the sample and the much larger explosiveness is observed at the end of the

sample, which is not compatible with our model (1). Nevertheless, the WLS method seems to detect the first explosiveness well, whereas  $\hat{k}_c$ , estimated by the OLS method, is no longer the collapsing date.

Figure 8 for etc in 2017 and Figure 9 for xmr in 2017 are similar to Figure 7 in that the time series has two explosiveness in the sample. Again, for etc in 2017, the first exuberance is well identified by the WLS method whereas the OLS method seems to fail to accurately estimate the recovering date. However, identifying the break dates by both methods for xmr in 2017 seems to be difficult.

Figure 10 shows the sample path of xlm in 2018, which has several small humps in this sample period. Although the three break dates estimated using the WLS may be interpreted as the emerging, collapsing, and recovering dates, they may not correspond to one specific explosiveness but to some of the several humps. On the other hand, the three estimated dates via the OLS method cannot be interpreted as designated by theory.

Figure 11 shows bnb in 2018, in which large explosiveness is observed at the beginning of the sample and a mild explosive and collapsing behavior seems to exist in most parts of the sample. It seems that the WLS method captures this second behavior, although the collapsing regime is relatively short when taking volatility shift into account. It seems that the estimated collapsing date by the OLS method seems to be incorrect and it may be either the recovering or emerging date.

As a whole, volatility correction based on the WLS method seems to work well, except for several cases where the explosive behavior is observed more than twice. We also observe that the WLS method can be robust to the short explosiveness either at the beginning or end of the sample if there exists another exuberance in the middle of the sample, although it is desirable to set up the sample periods in which only one set of the exuberance is included. For that purpose, the procedure proposed by Phillips et al. (2015a,b) may be useful.

## 7 Conclusion

We propose an algorithm for volatility correction in the estimation of the dates of the bubble in the four-regime model. The method consists of the following steps: estimation of the break dates without volatility correction and calculation of the residuals, and WLS-based estimation of the dates of the bubble using the weight function obtained through the non-parametric estimation of the volatility function. The Monte Carlo results show that the estimated break dates are at least as accurate as those under the homoskedasticity assumption and

better in some cases. The empirical illustration using the cryptocurrencies demonstrates the different performance of the two methods, with and without volatility correction, while also establishing that the WLS method returns the adequate break dates more often than the OLS method.

## References

- Astill, S., Harvey, D. I., Leybourne, S. J., Taylor, A. M. R., and Zu, Y. (2023). CUSUM-based monitoring for explosive episodes in financial data in the presence of time-varying volatility. *Journal of Financial Econometrics*, 21:187–227.
- Cavaliere, G. and Taylor, A. M. R. (2007a). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, 140(2):919–947.
- Cavaliere, G. and Taylor, A. M. R. (2007b). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis*, 29(2):300–330.
- Harris, D., Kew, H., and Taylor, A. M. R. (2020). Level shift estimation in the presence of non-stationary volatility with an application to the unit root testing problem. *Journal of Econometrics*, 219(2):354–388.
- Harvey, D. I., Leybourne, S. J., and Sollis, R. (2017). Improving the accuracy of asset price bubble start and end date estimators. *Journal of Empirical Finance*, 40:121–138.
- Harvey, D. I., Leybourne, S. J., Sollis, R., and Taylor, A. M. R. (2016). Tests for explosive financial bubbles in the presence of non-stationary volatility. *Journal of Empirical Finance*, 38:548–574.
- Harvey, D. I., Leybourne, S. J., and Zu, Y. (2018). Testing explosive bubbles with time-varying volatility. *Econometric Reviews*, 38:1131–1151.
- Harvey, D. I., Leybourne, S. J., and Zu, Y. (2020). Sign-based unit root tests for explosive financial bubbles in the presence of deterministically time-varying volatility. *Econometric Theory*, 36(1):122–169.
- Kurozumi, E. and Skrobotov, A. (2022). On the asymptotic behavior of bubble date estimators. *Journal of Time Series Analysis*, 44:359–373.
- Kurozumi, E., Skrobotov, A., and Tsarev, A. (2023). Time-transformed test for bubbles under non-stationary volatility. *Journal of Financial Econometrics*, 21:1282–1307.

- Pang, T., Du, L., and Chong, T. T.-L. (2021). Estimating multiple breaks in nonstationary autoregressive models. *Journal of Econometrics*, 221(1):277–311.
- Phillips, P. C. B. and Shi, S. (2020). Real time monitoring of asset markets: Bubbles and crises. In *Handbook of Statistics*, volume 42, pages 61–80. Elsevier.
- Phillips, P. C. B., Shi, S., and Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4):1043–1078.
- Phillips, P. C. B., Shi, S., and Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review*, 56(4):1079–1134.
- Phillips, P. C. B., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review*, 52(1):201–226.
- Xu, K.-L. and Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics*, 142(1):265–280.

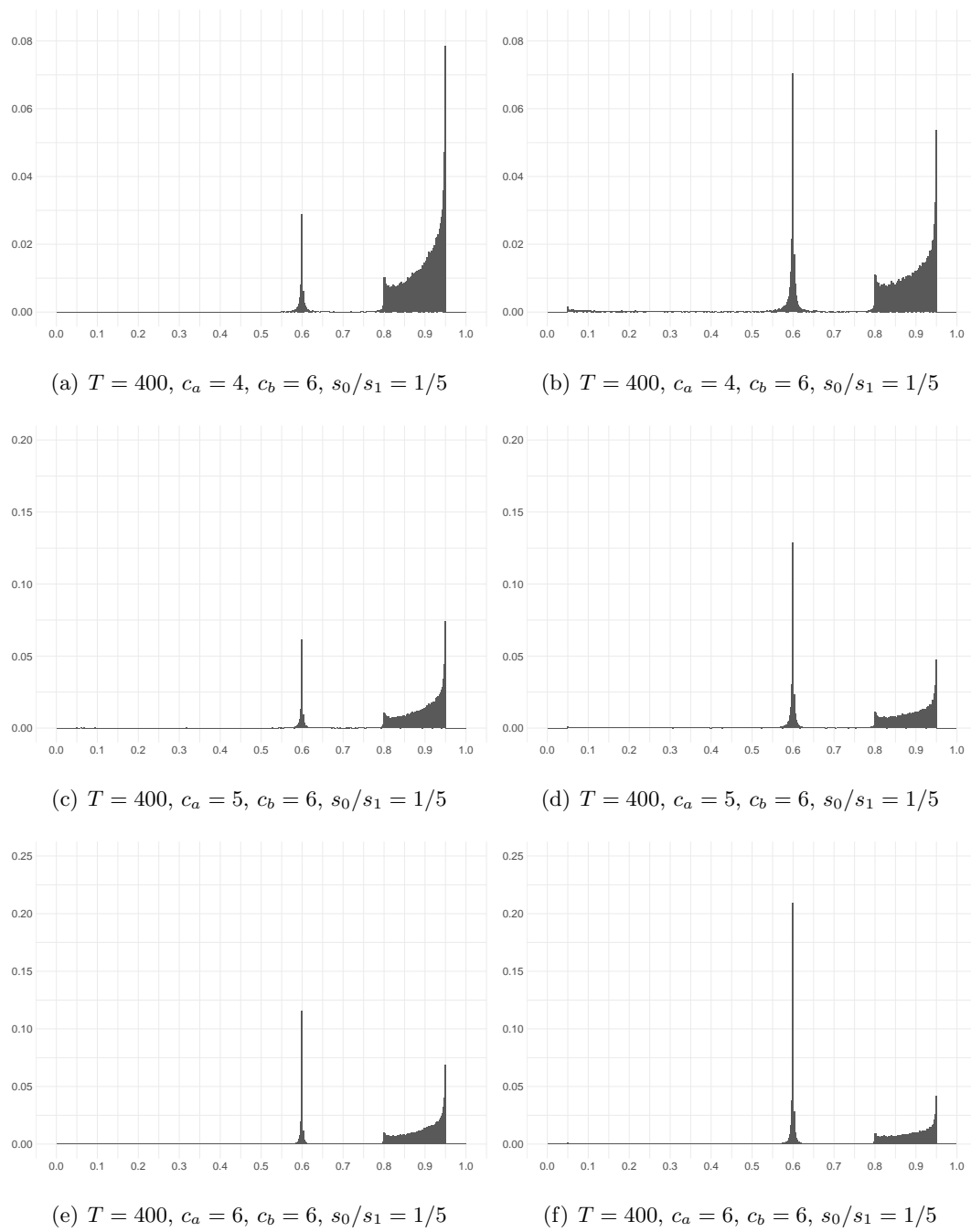


Figure 1: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$

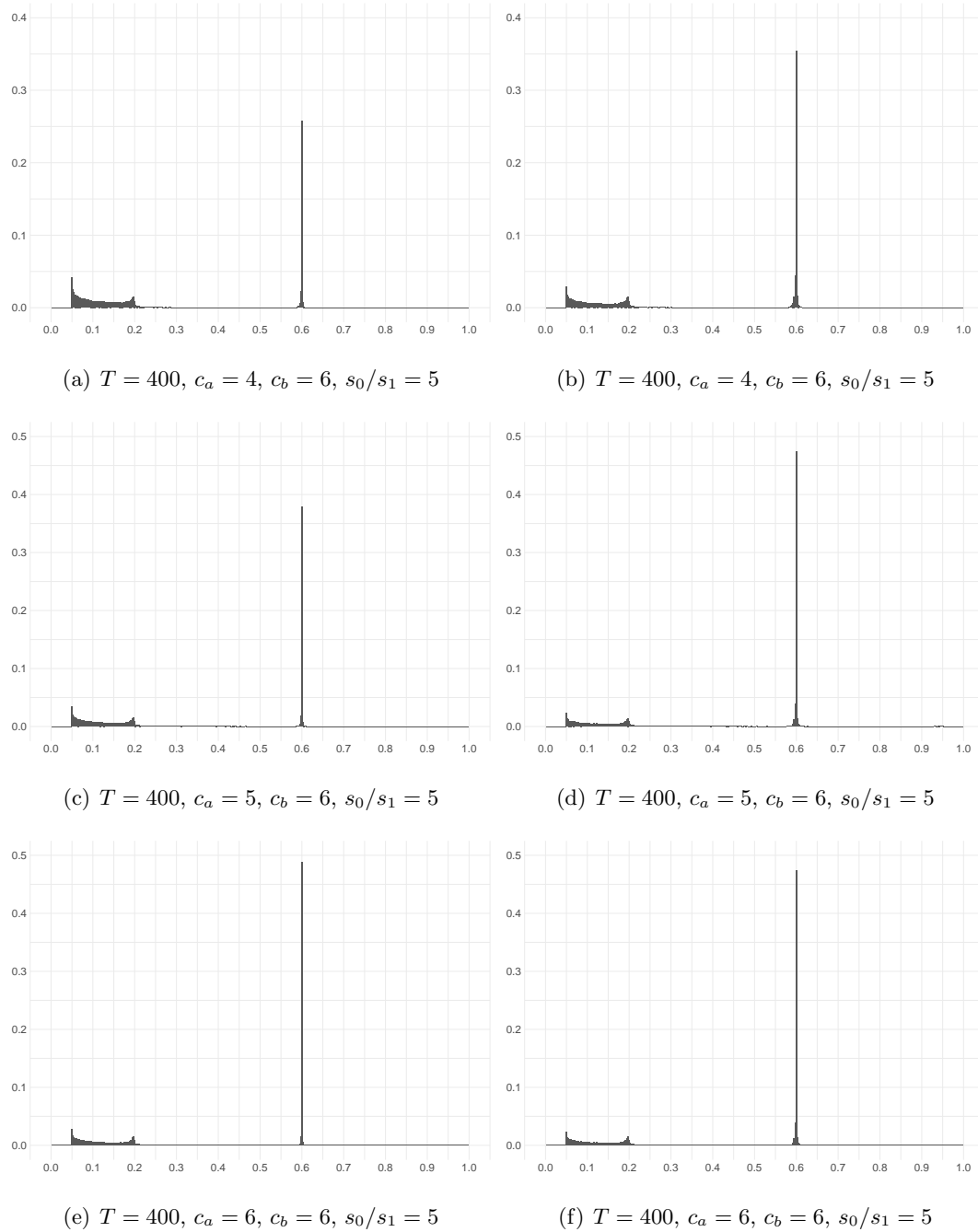


Figure 2: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 400$

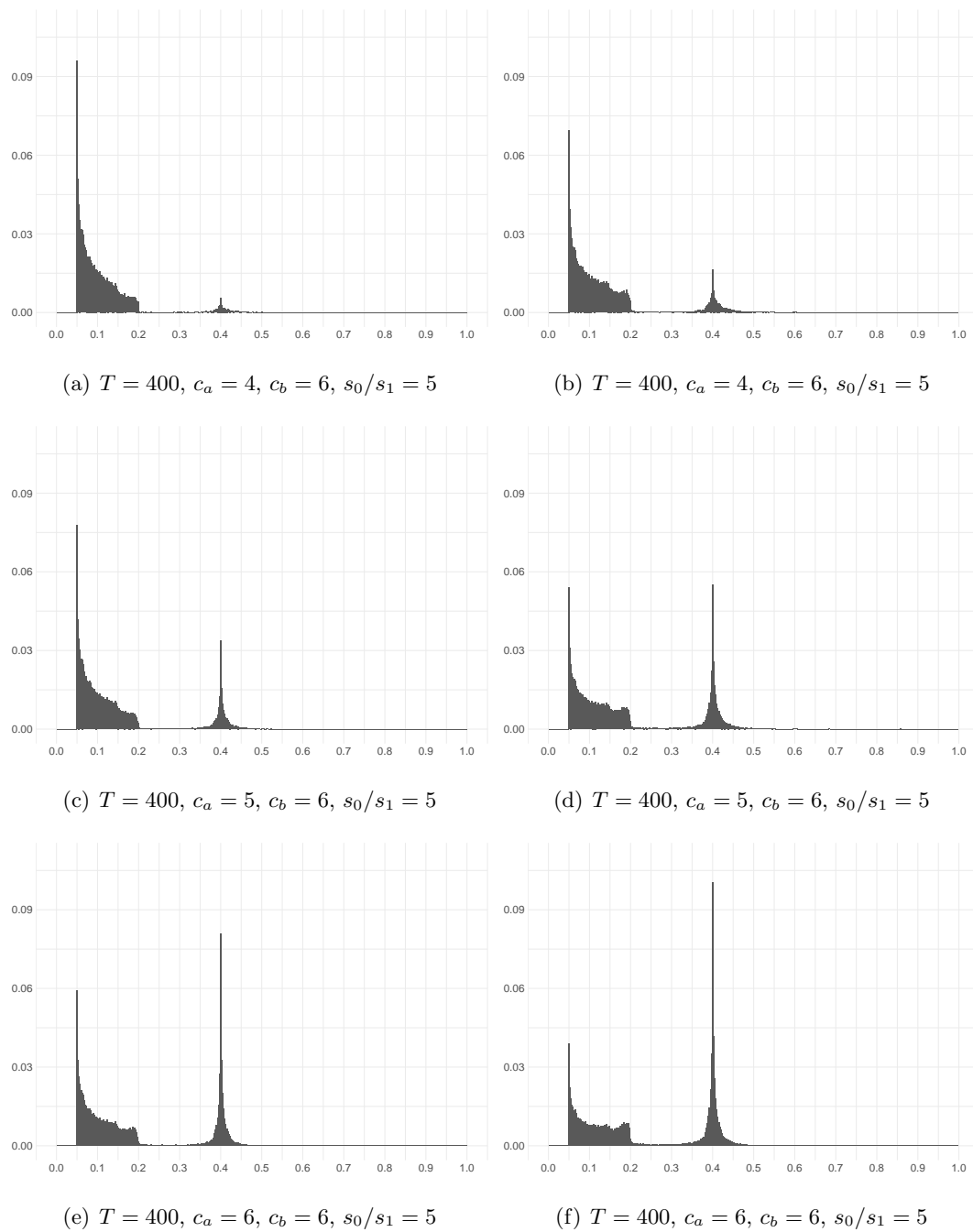


Figure 3: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 400$

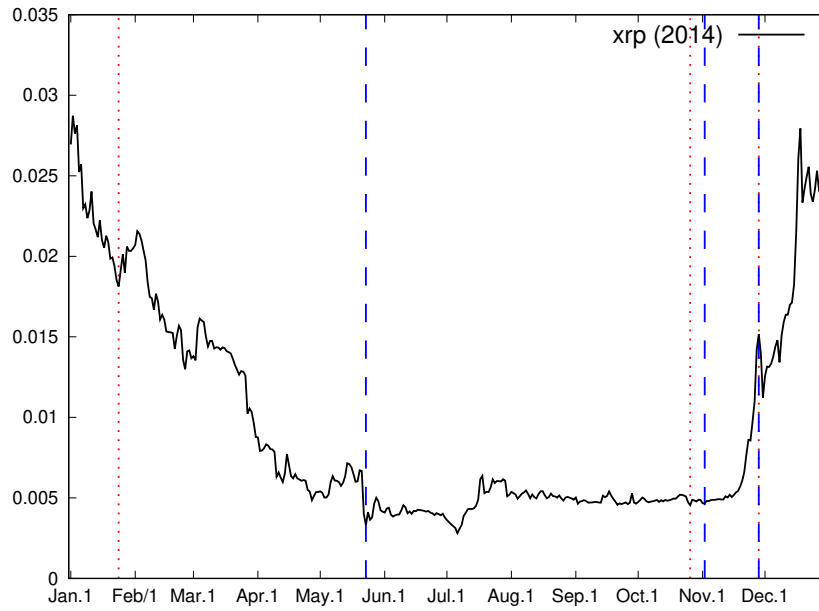


Figure 4: XRP in 2014

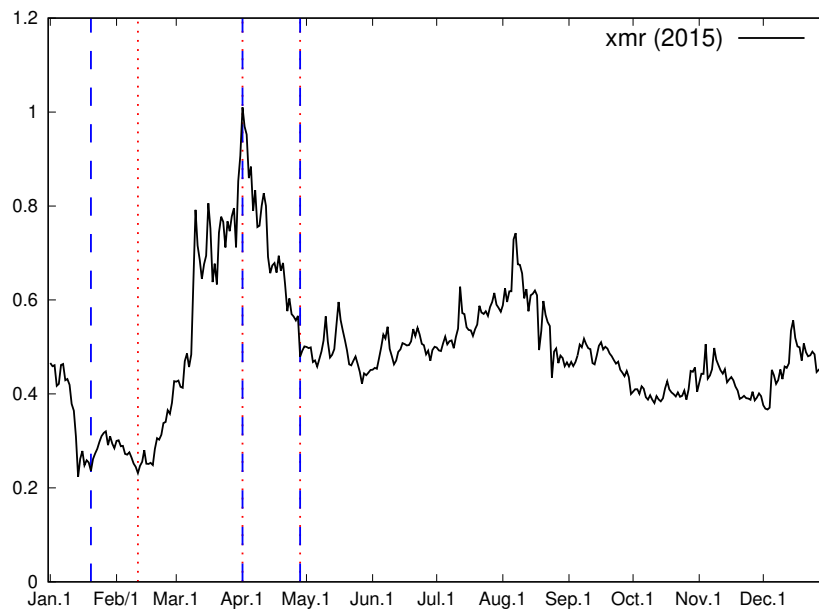


Figure 5: XMR in 2015





Figure 6: ETH in 2016

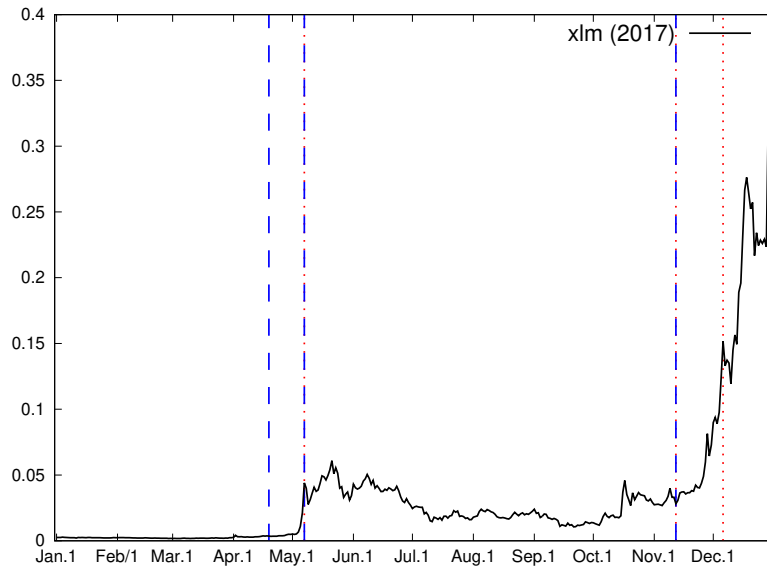


Figure 7: XLM in 2017

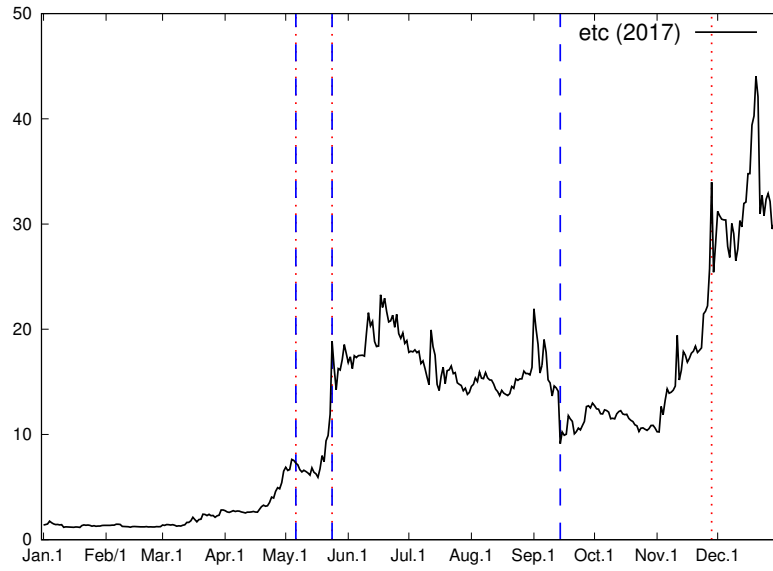


Figure 8: ETC in 2017

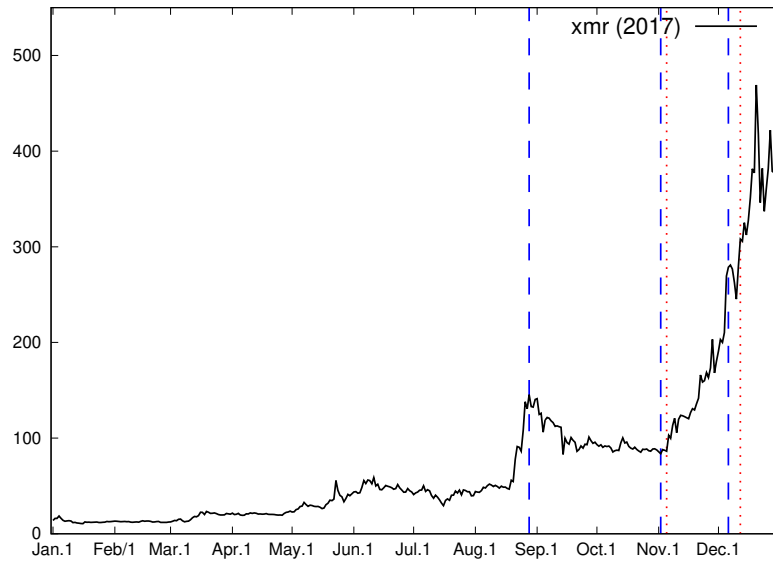


Figure 9: XMR in 2017

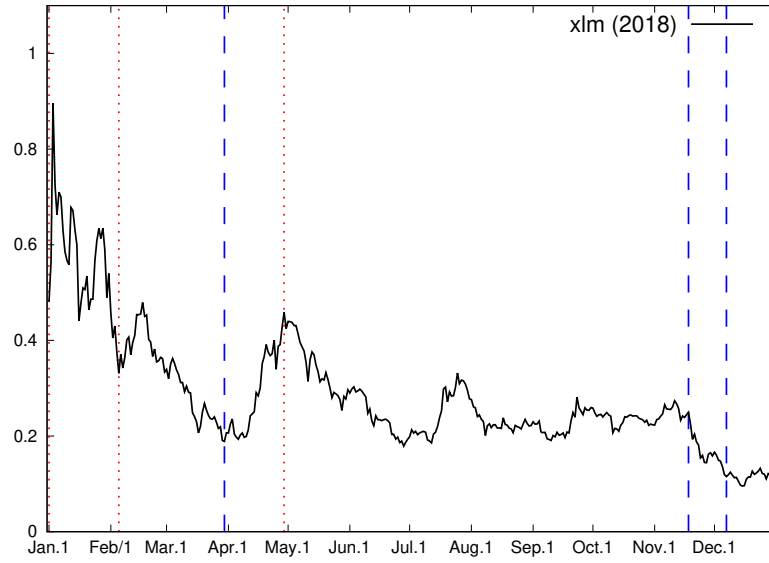


Figure 10: XLM



Figure 11: BNB in 2018

**Online Appendix to**  
**“Improving the accuracy of bubble date estimators under**  
**time-varying volatility”**  
**by Eiji Kurozumi and Anton Skrobotov**

This appendix contains additional figures with the histograms of the break date estimates with different combinations of  $s_0$ ,  $s_1$ ,  $\tau$ , and  $T$ , as described in Section 5 of the main part of the paper. Figures A.1–C.6 correspond to the case where the volatility shift occurs late in the sample at  $\tau = 0.8$  (Figures A.1–A.6 for  $\hat{k}_c$ , B.1–B.6 for  $\hat{k}_e$ , and C.1–C.6 for  $\hat{k}_r$ ), while Figures D.1–F.6 are the histograms in the case where the volatility changes early in the sample at  $\tau = 0.2$  (Figures D.1–D.6 for  $\hat{k}_c$ , E.1–E.6 for  $\hat{k}_e$ , and F.1–F.6 for  $\hat{k}_r$ ).

Figure A.1 ( $\hat{k}_c$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$ ) is the same as Figure 1 in the main text and Figure A.2 with  $T = 800$  is quantitatively similar to Figure A.1; the sample splitting approach based on the WLS method performs better than that based on the OLS method. Figures A.3–A.6 show that there is virtually no difference between the distributions of the estimates with and without volatility correction.

Figures B.1–B.2 ( $\hat{k}_e$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ) demonstrate the local peak of  $\hat{\tau}_e$  at an incorrect location (at 0.6, the location of collapse), but volatility correction reduces this local peak and increases the peak at 0.4, the correct location of bubble exuberance. For the cases where  $s_0/s_1 = 1$  and  $s_0/s_1 = 5$  in Figures B.3–B.6, the results are virtually the same regardless of the correction.

Figures C.1–C.2 ( $\hat{k}_r$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ) demonstrate the local peak at an incorrect location (at the end of the sample), but volatility correction reduces this local peak and increases the peak at the correct location of the recovering date at  $\tau_r = 0.7$ . Figures C.3–C.6 show no significant difference between the two methods.

For  $\tau = 0.2$  and  $\hat{k}_c$ , we observe the difference between the with and without correction scenario only for the case of  $s_0/s_1 = 5$  (Figures D.5 and D.6, the former of which is the same as Figure 2 in the main text). Figures D.1–D.4 show similar results regardless of whether we correct for volatility or not.

For  $\hat{k}_e$ , it is difficult to interpret the performance of the estimates for  $s_0/s_1 = 1/5$  (Figures E.1–E.2). For  $s_0/s_1 = 1$ , the results are virtually the same regardless of whether we correct for the volatility or not (Figures E.3–E.4), whereas for  $s_0/s_1 = 5$ , the local peak becomes higher around the true break fraction under volatility correction, as is observed in Figures E.5 and E.6.

For  $\hat{k}_r$  and  $s_0/s_1 = 1/5$ , Figures F.1–F.2 demonstrate the local peak at an incorrect location (at the end of the sample), but volatility correction reduces this local peak and increases the peak at the correct location of bubble exuberance. For the case  $s_0/s_1 = 1$ , the results are similar regardless of whether we correct for the volatility or not (Figures F.3–F.4). For the case  $s_0/s_1 = 5$ , volatility correction reduces the local peak at the incorrect location (corresponding to the date of collapse) and increases the peak at the correct location of resumption of the normal market (Figures F.5–F.6).

A.  $\tau = 0.8, \hat{k}_c$

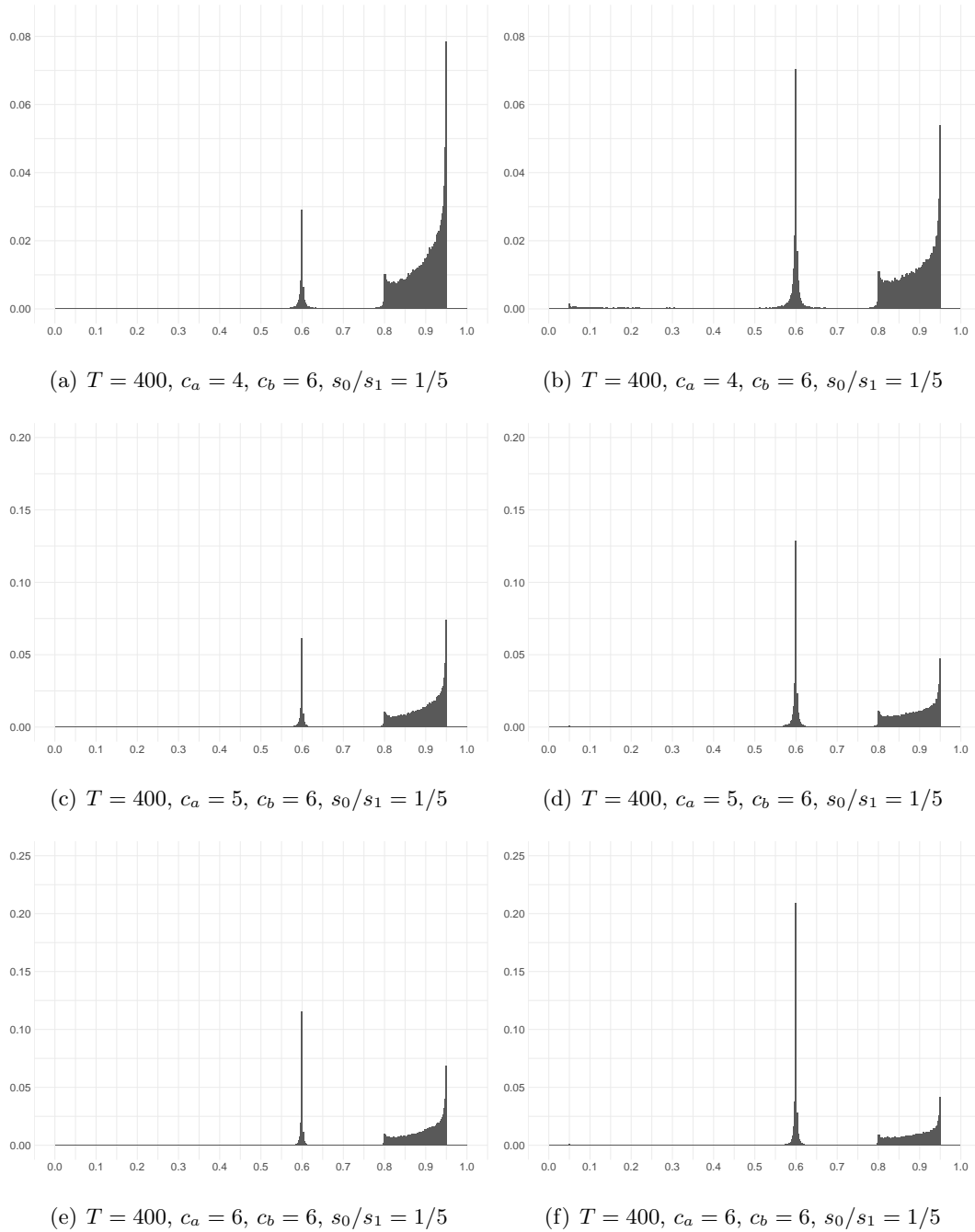


Figure A.1: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$

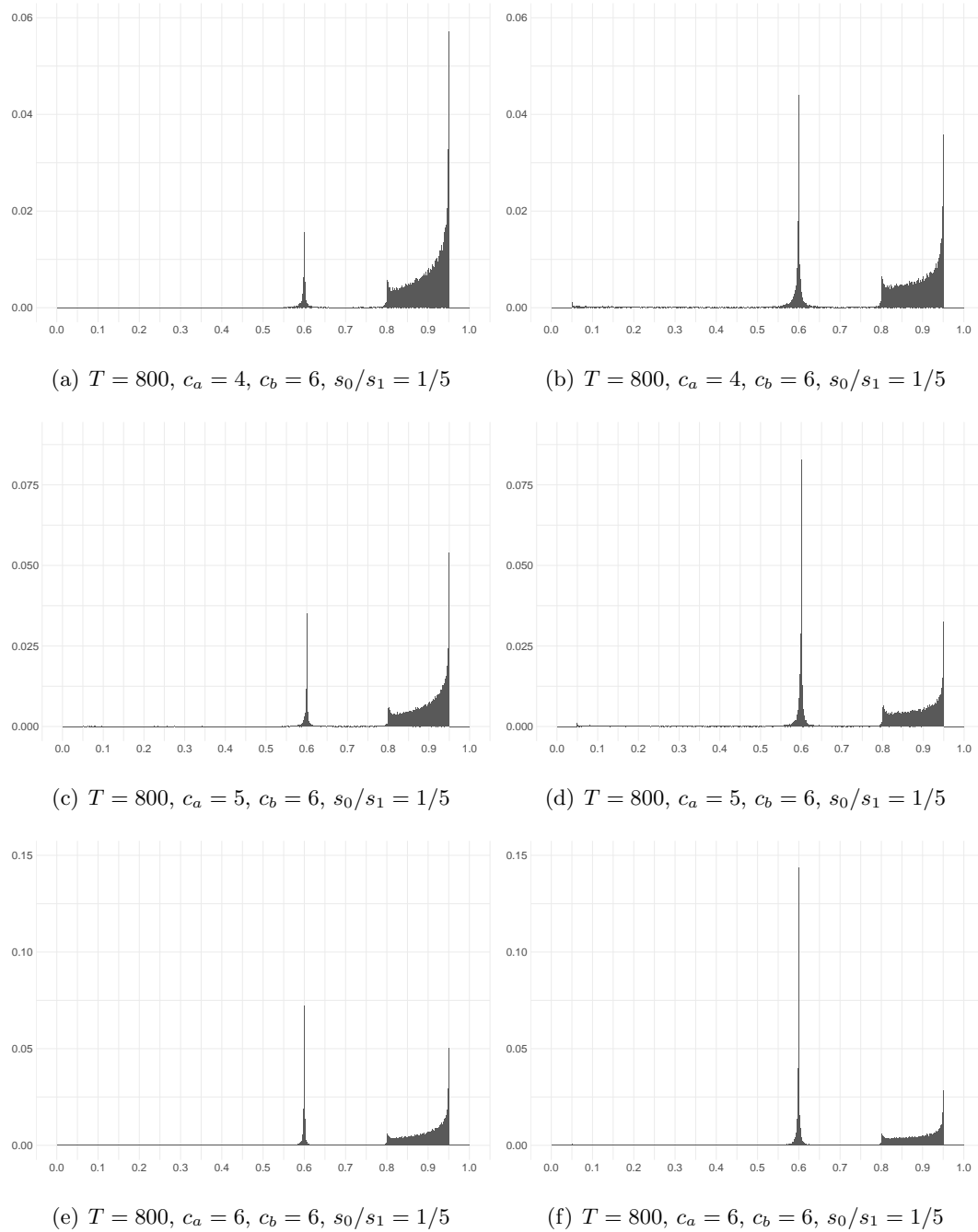


Figure A.2: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 800$

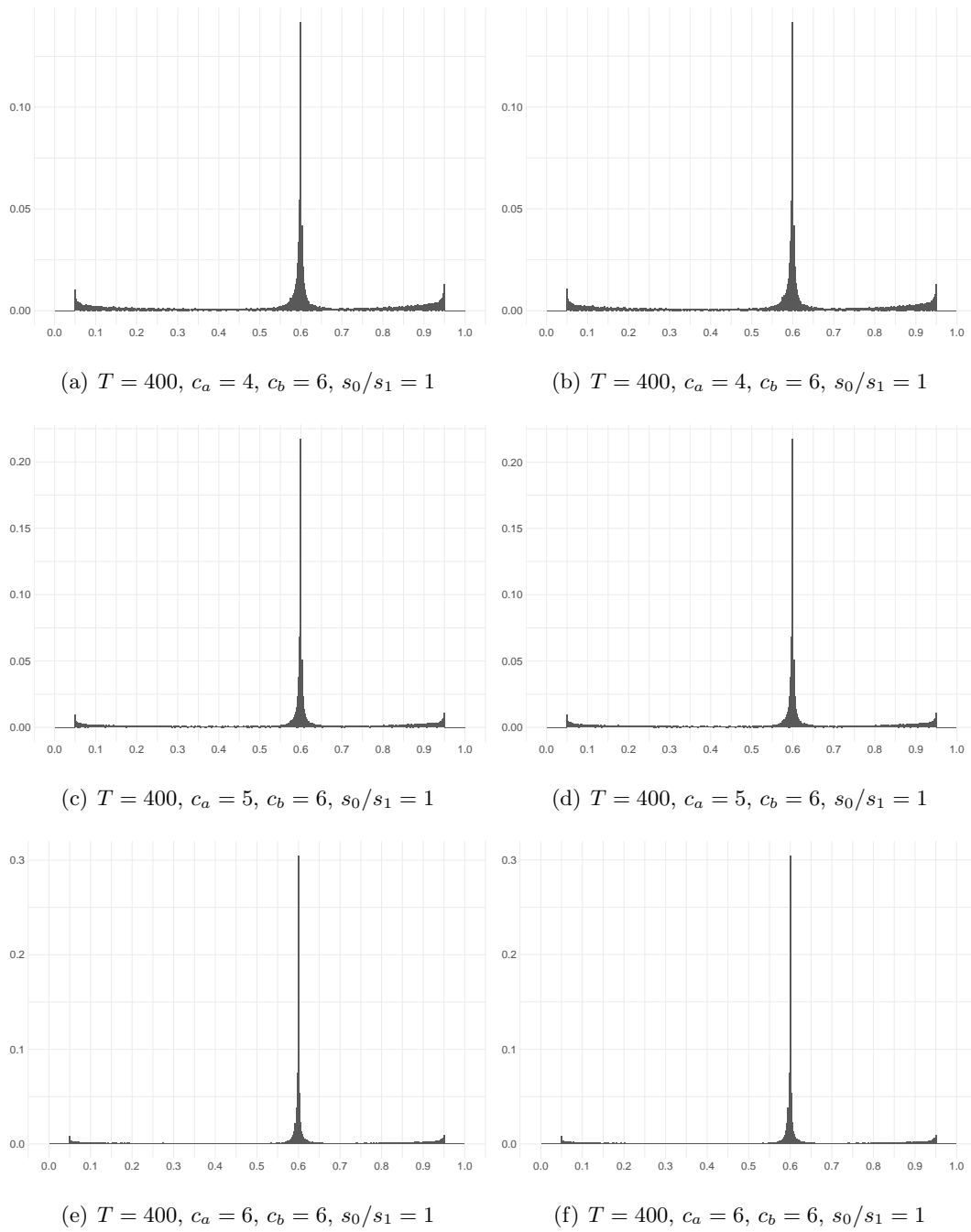


Figure A.3: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1$ ,  $T = 400$



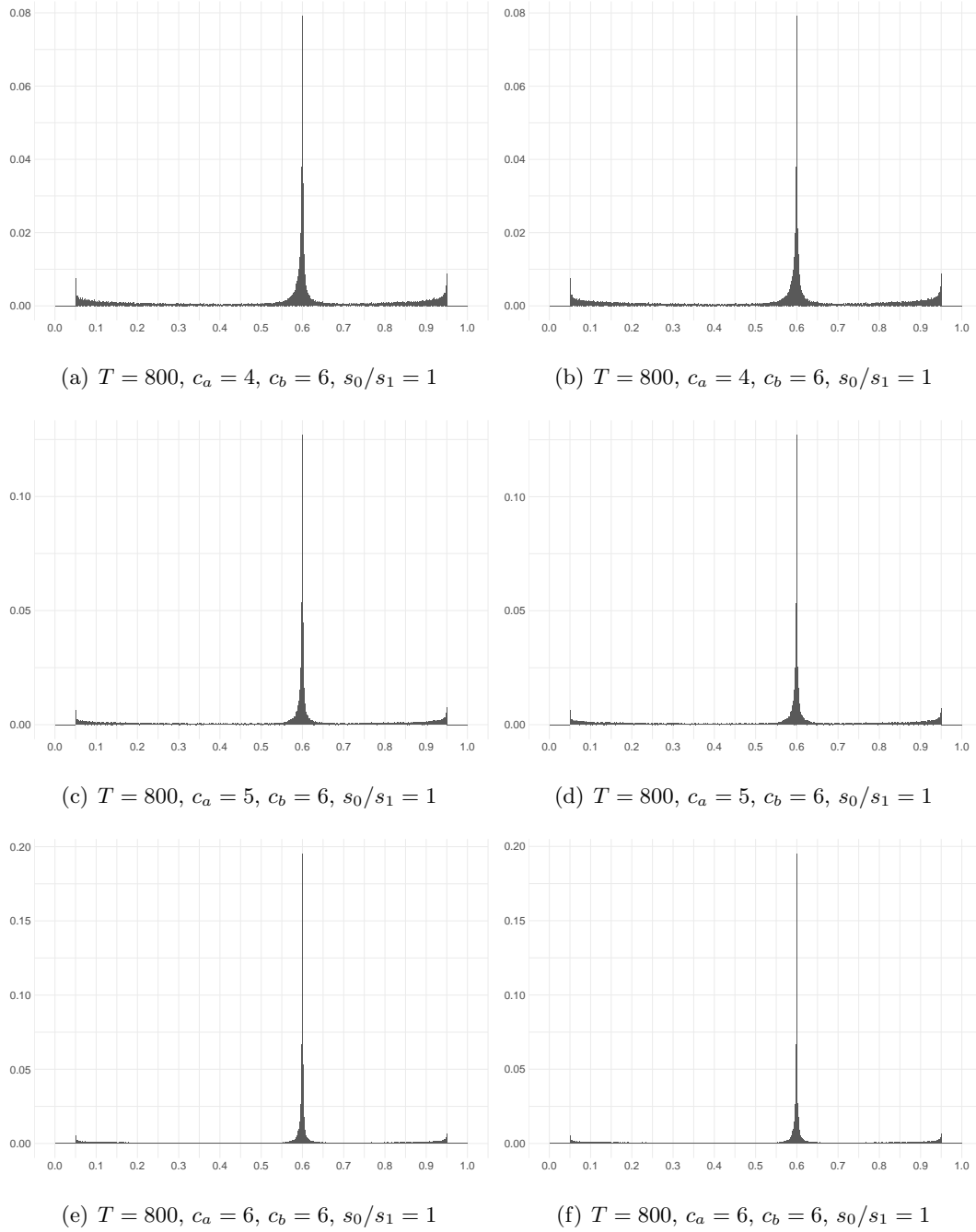


Figure A.4: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1$ ,  $T = 800$

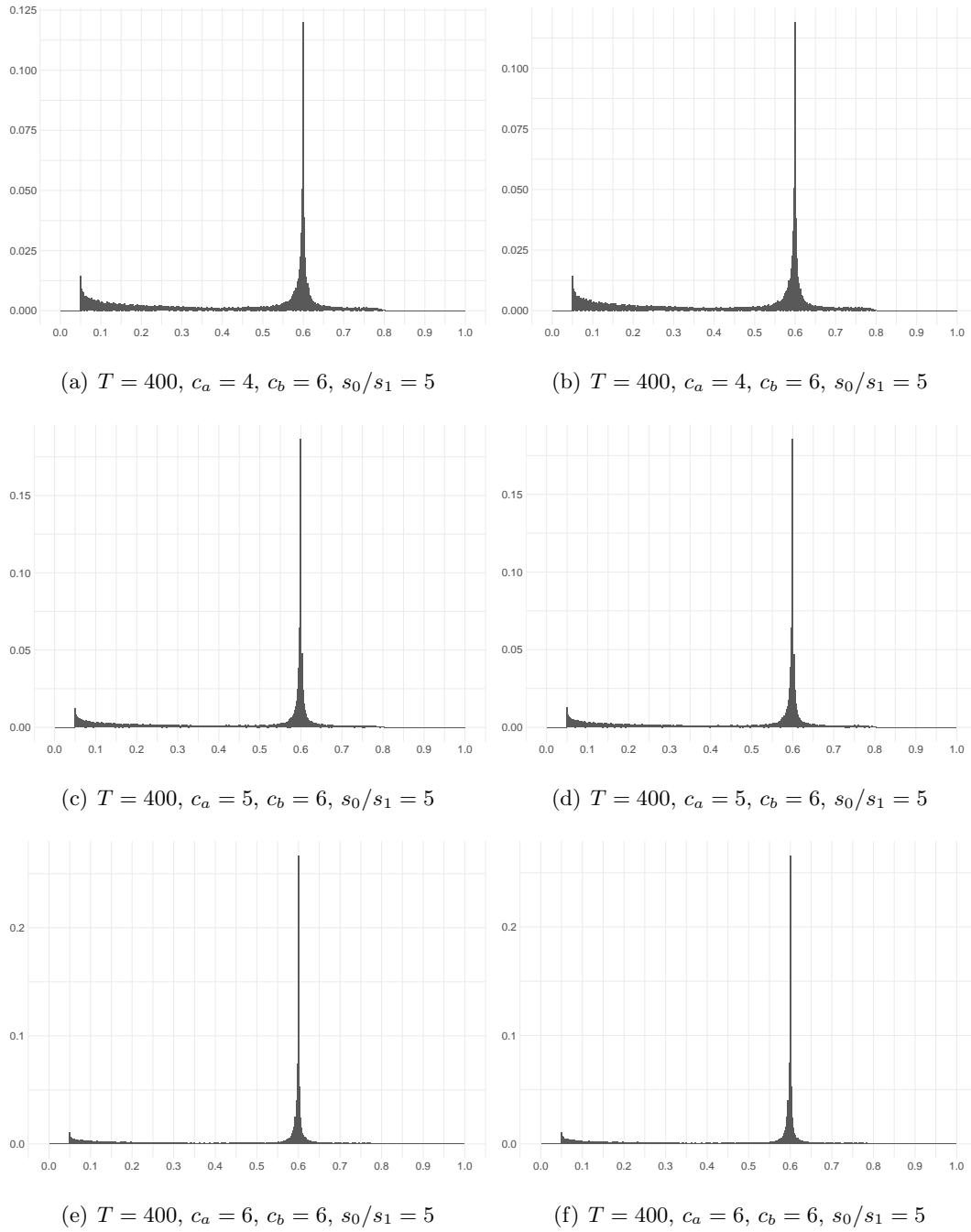


Figure A.5: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 5$ ,  $T = 400$

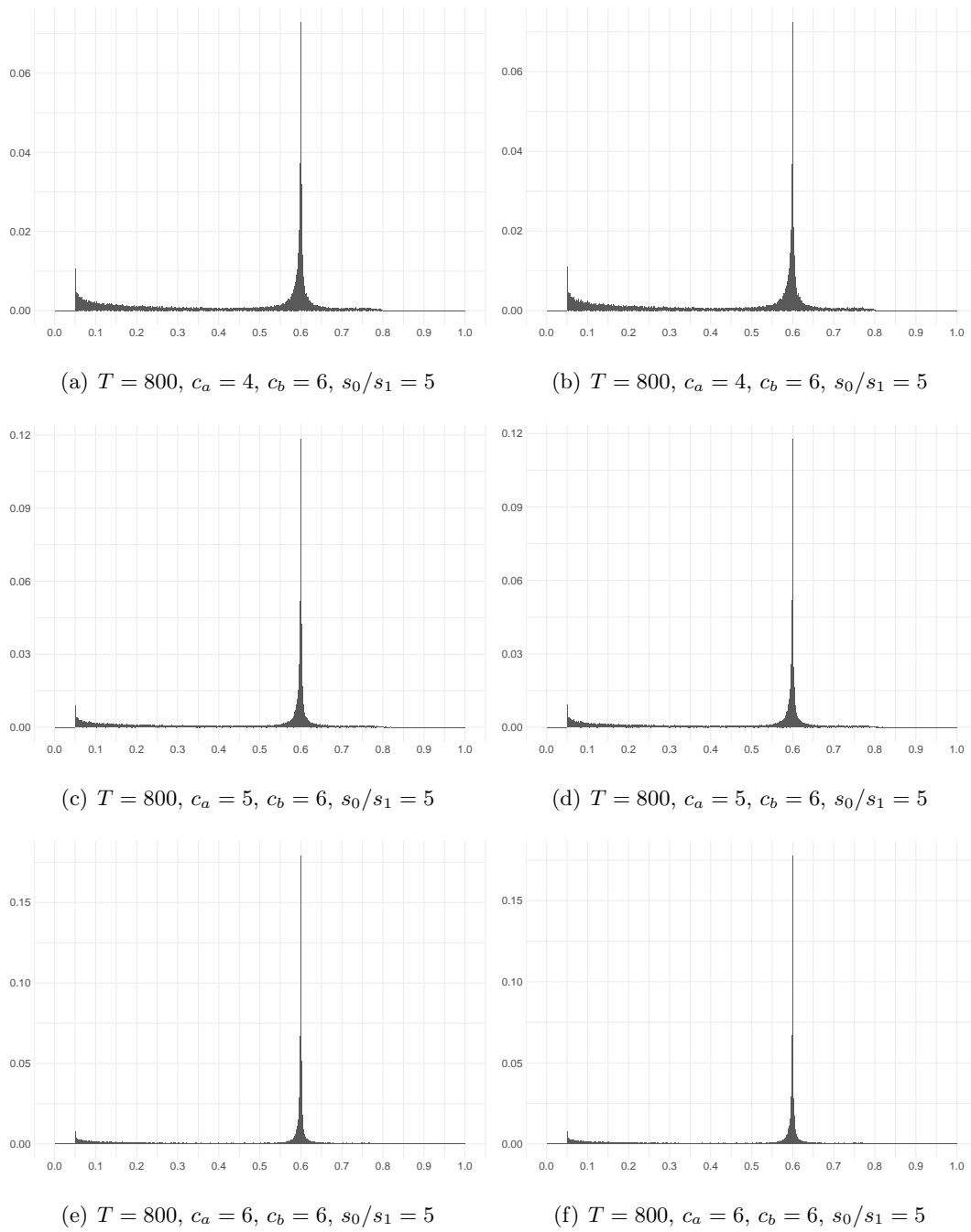


Figure A.6: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 5$ ,  $T = 800$

**B.**  $\tau = 0.8, \hat{k}_e$

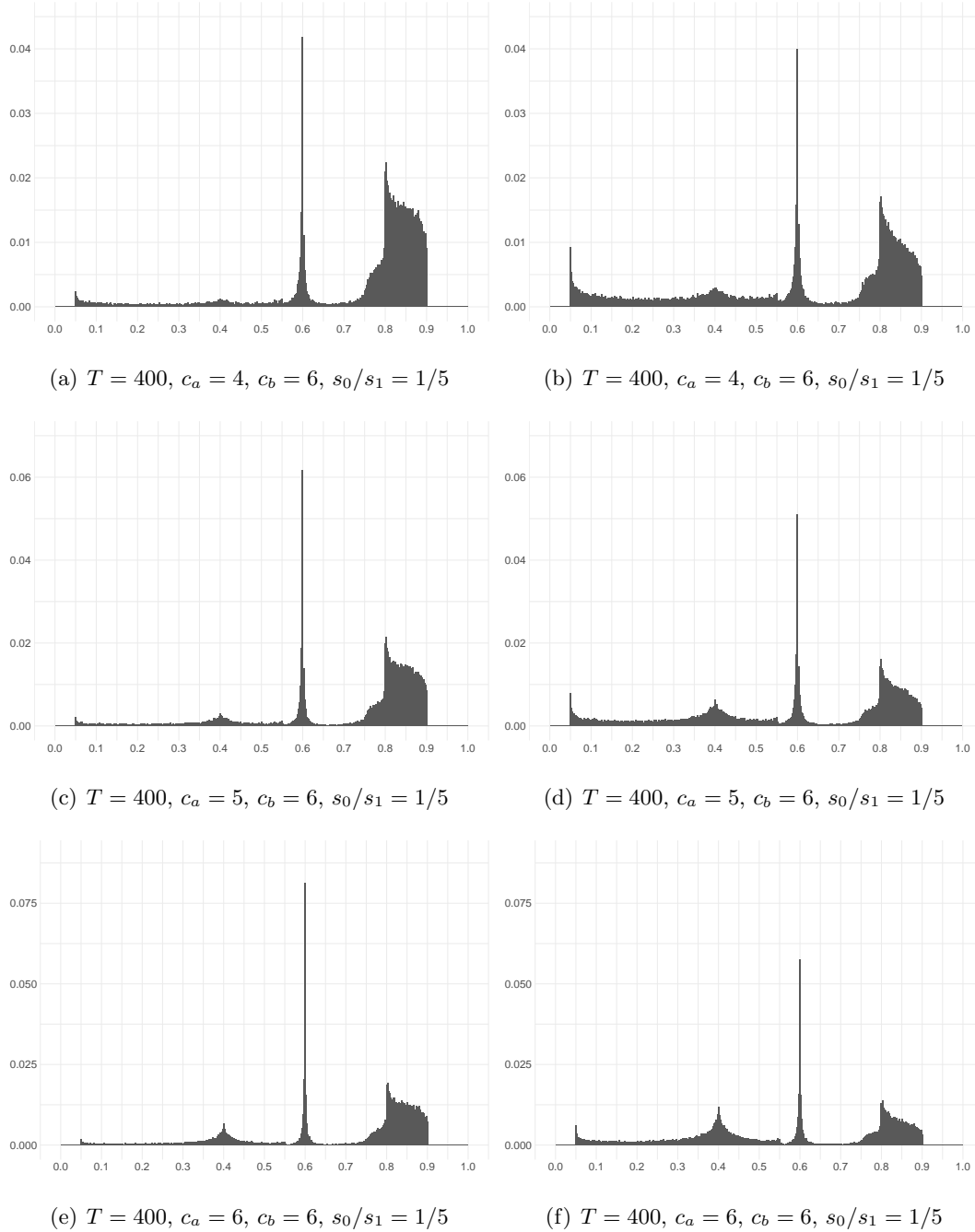


Figure B.1: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$

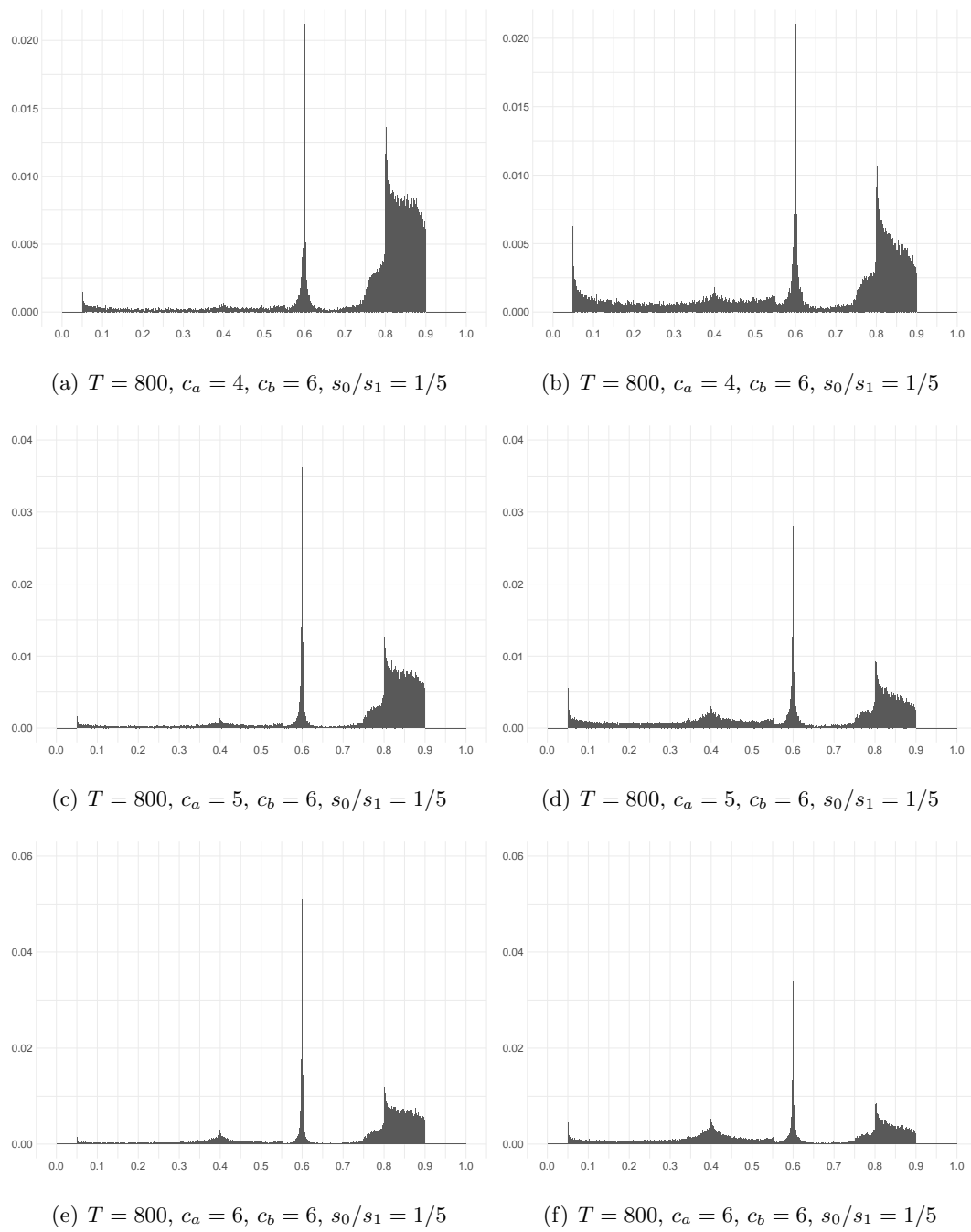


Figure B.2: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 800$

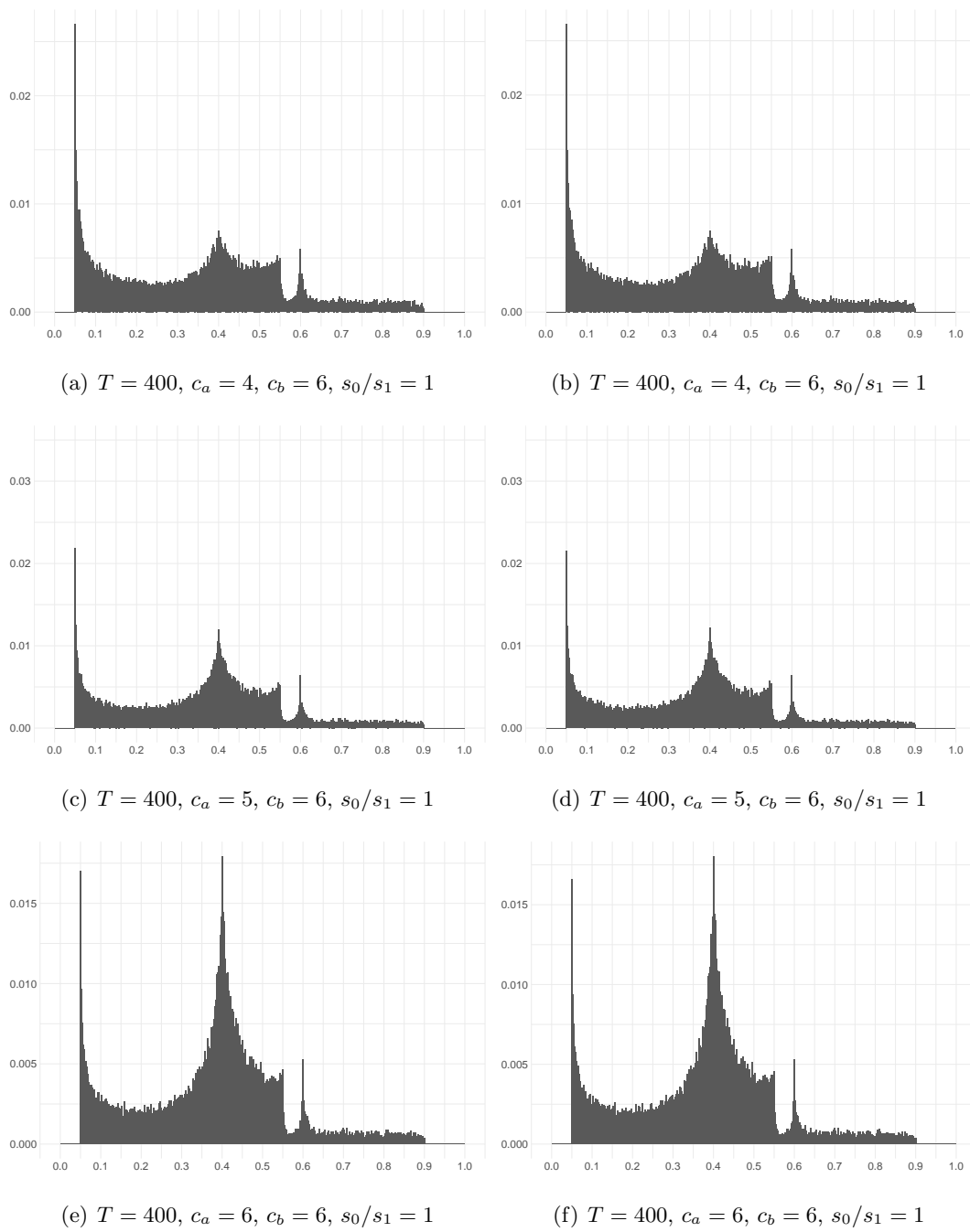


Figure B.3: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1$ ,  $T = 400$

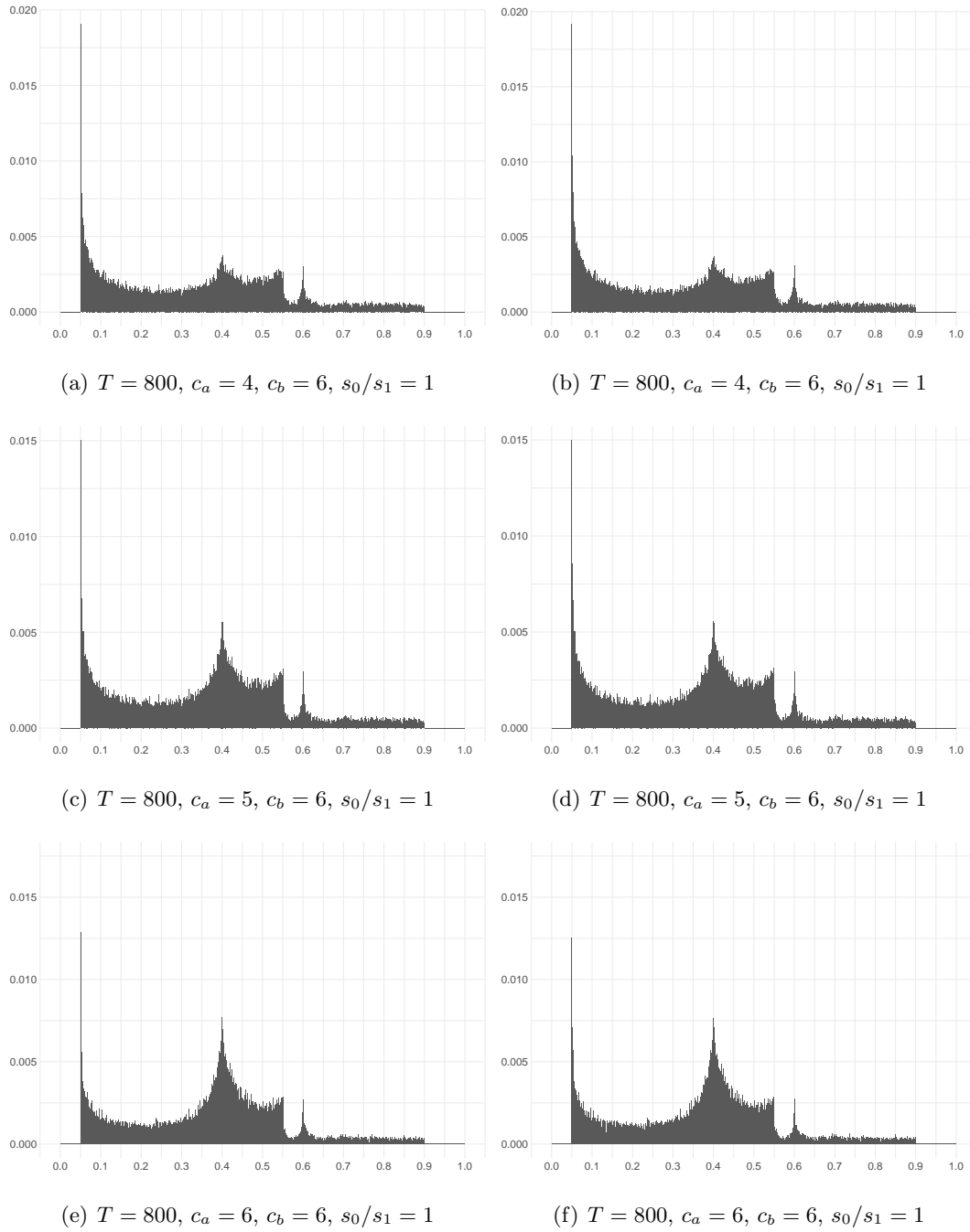


Figure B.4: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1$ ,  $T = 800$

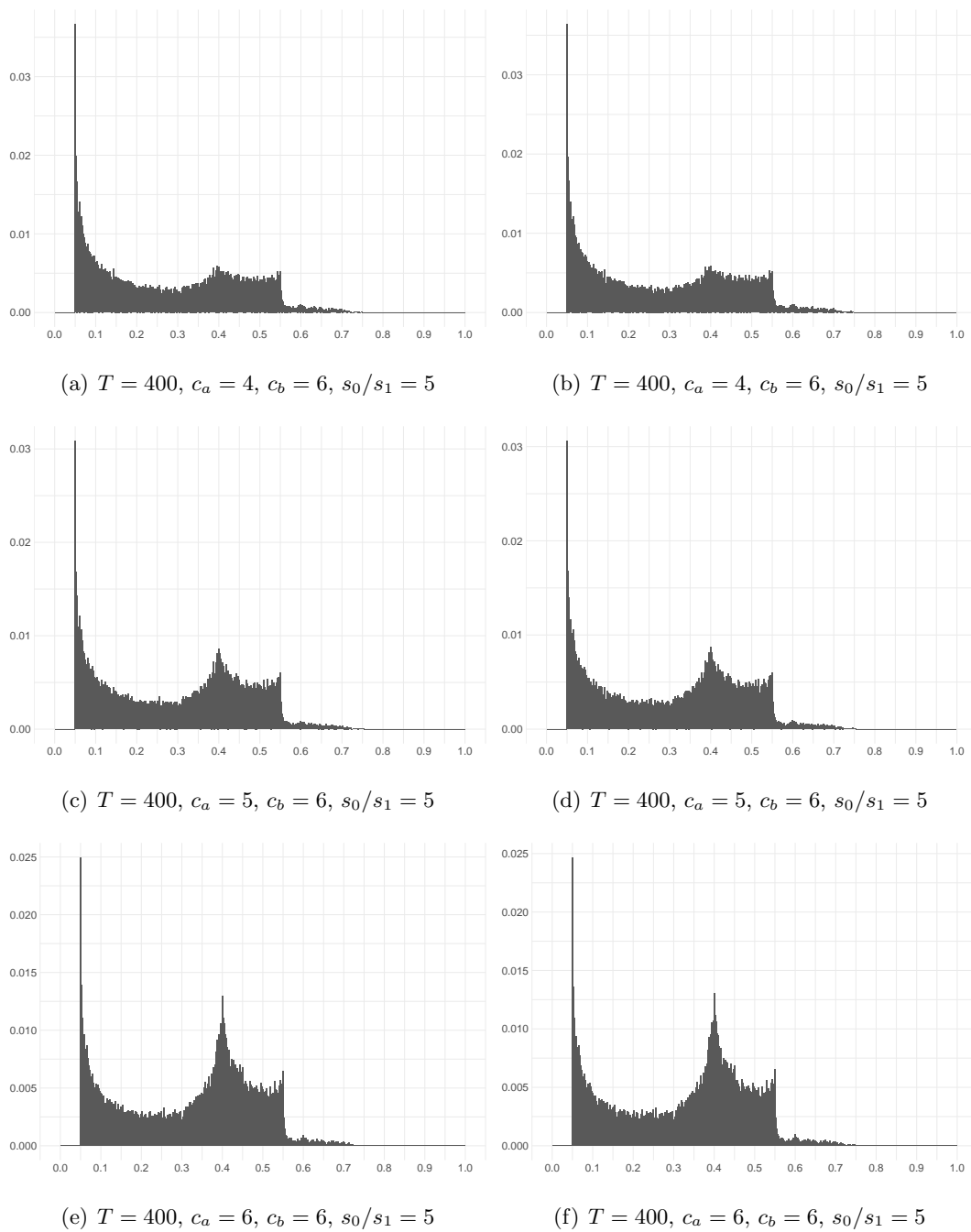


Figure B.5: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 5$ ,  $T = 400$



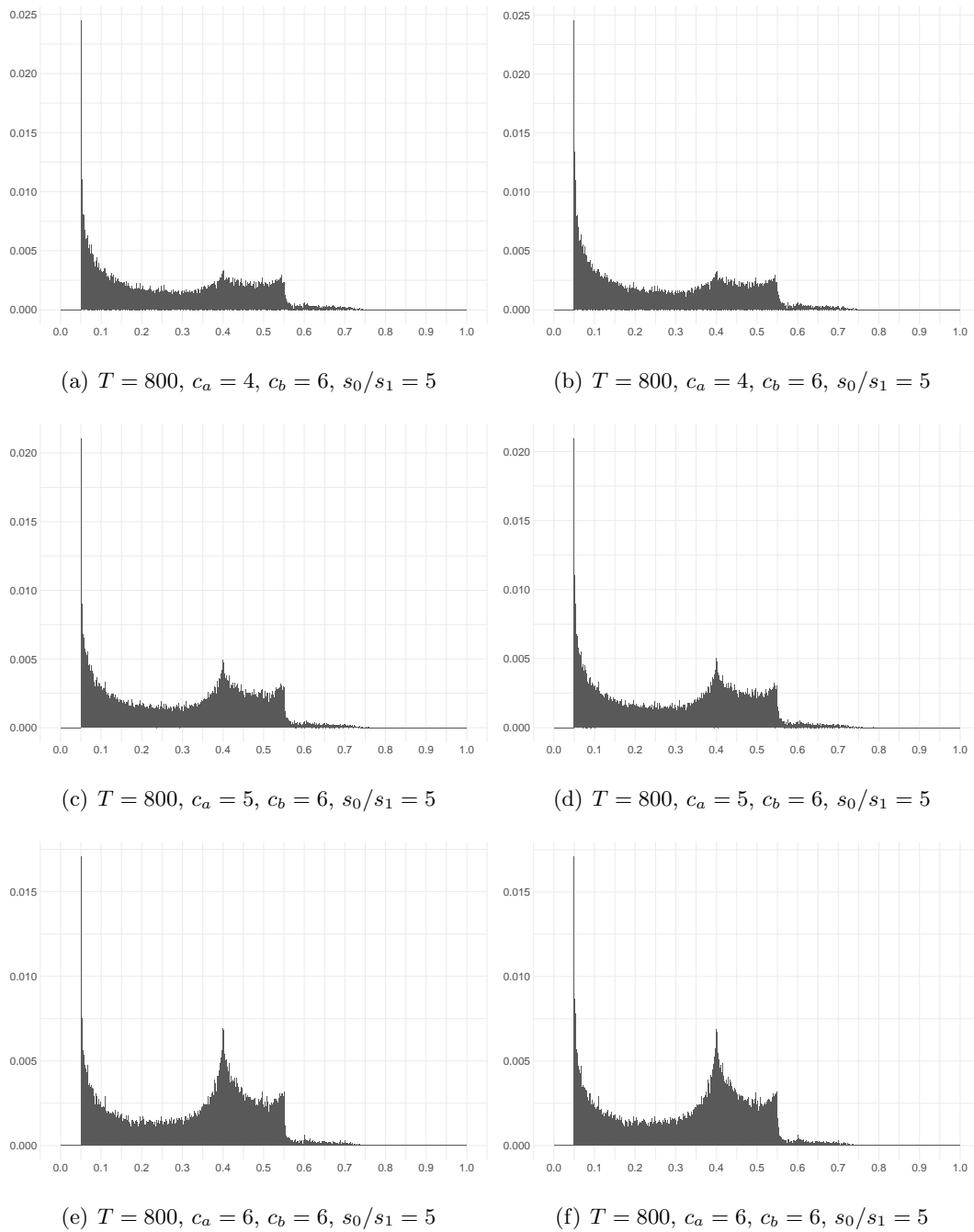


Figure B.6: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 5$ ,  $T = 800$

C.  $\tau = 0.8, \hat{k}_r$

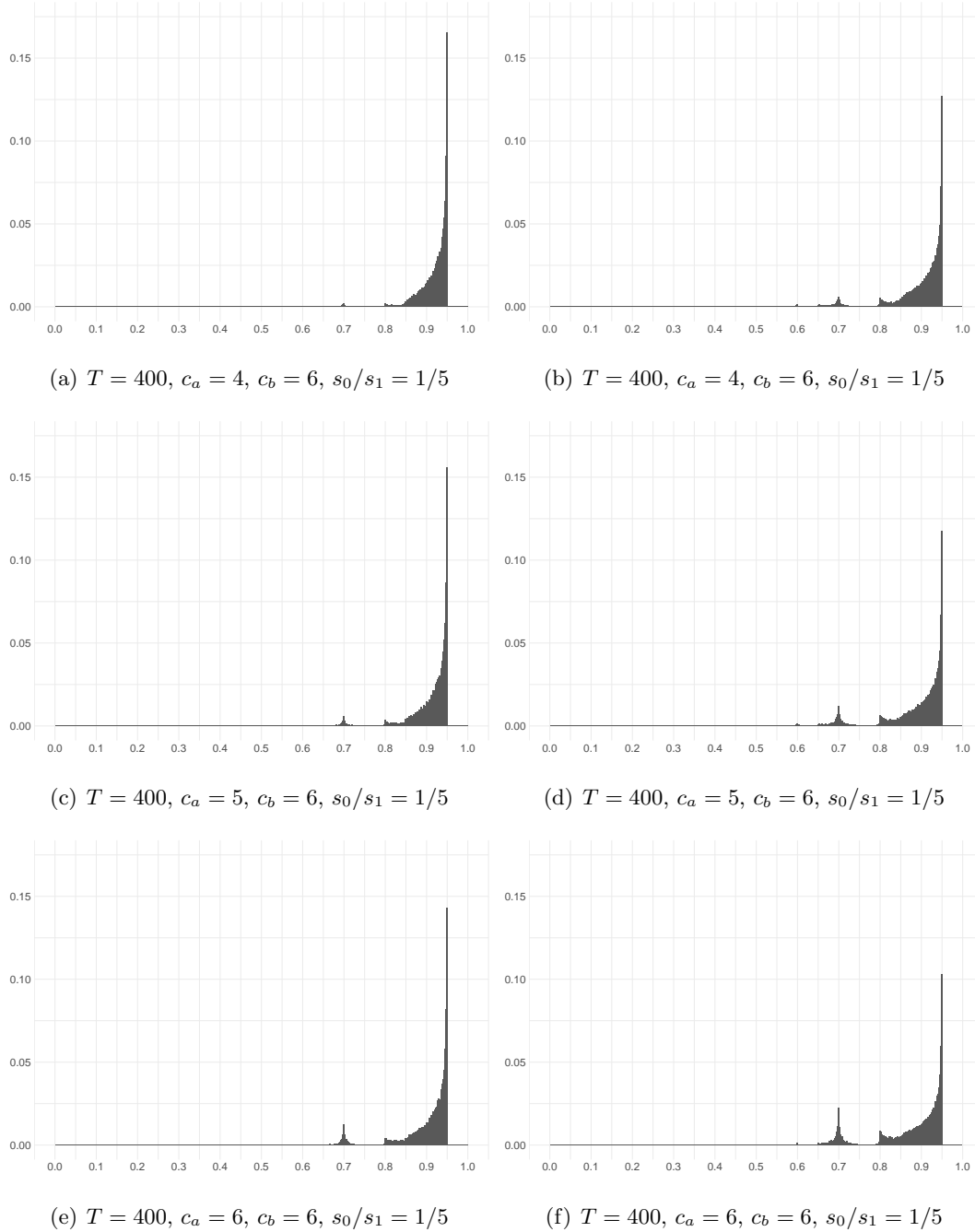


Figure C.1: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$

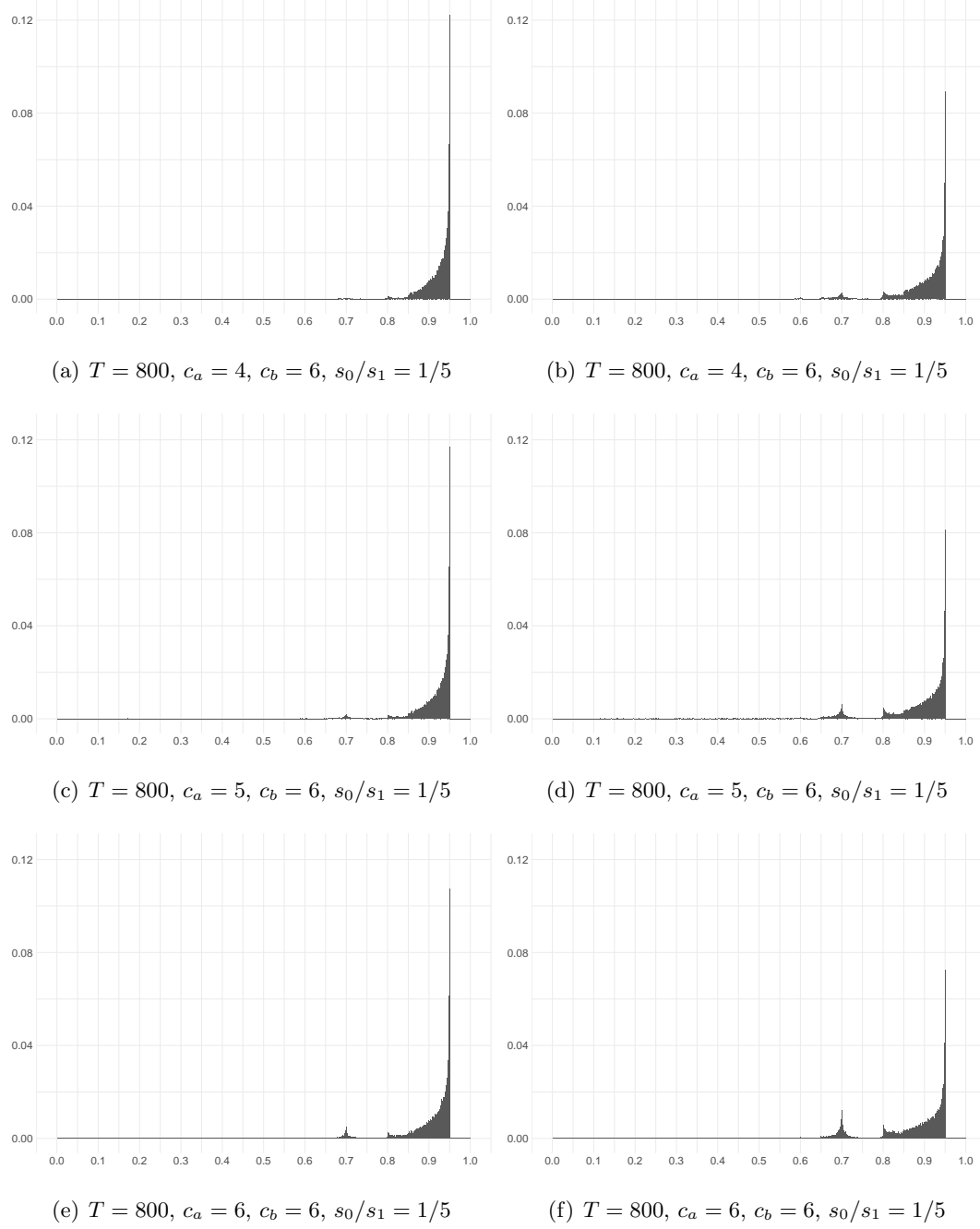


Figure C.2: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1/5$ ,  $T = 800$

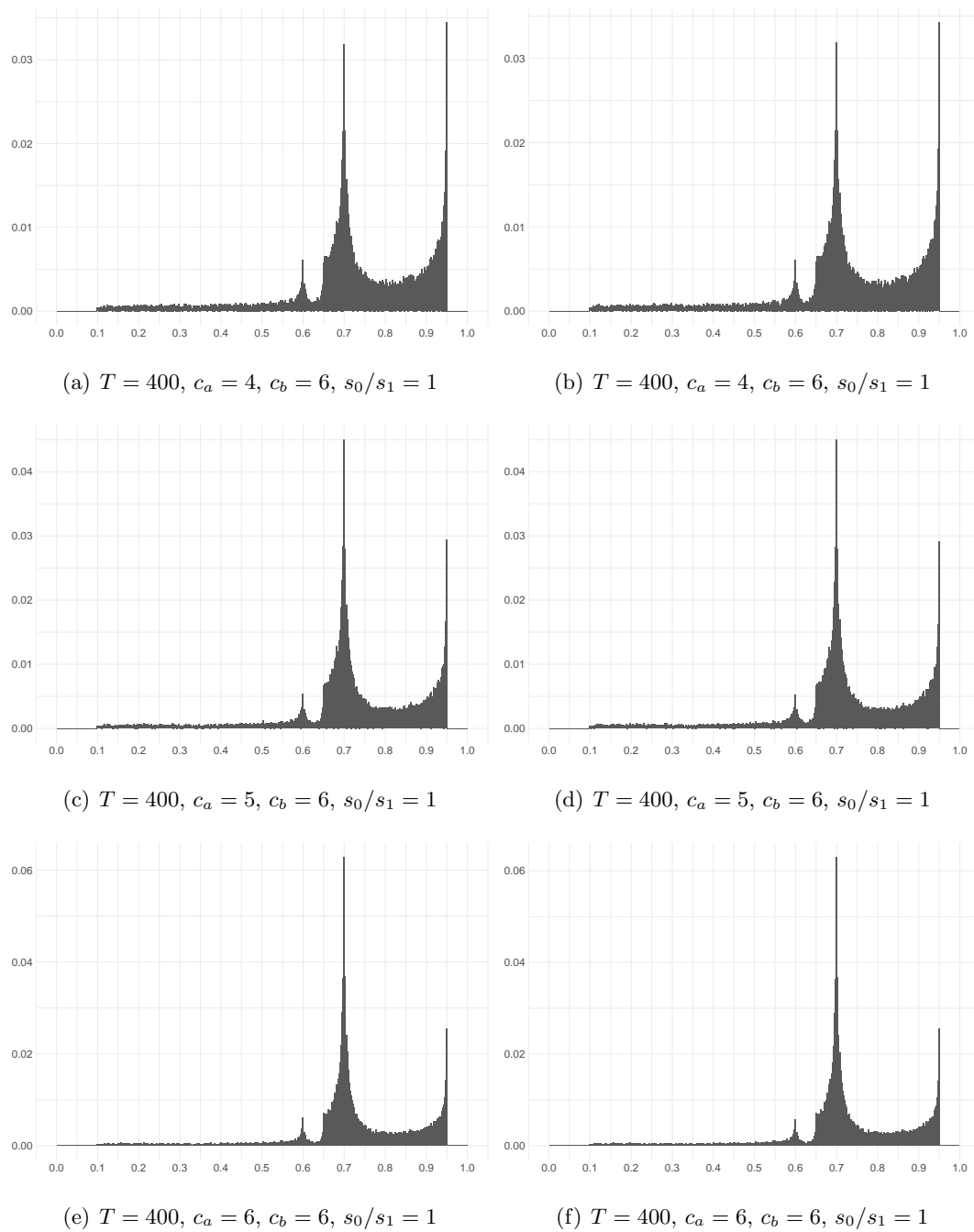


Figure C.3: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1$ ,  $T = 400$

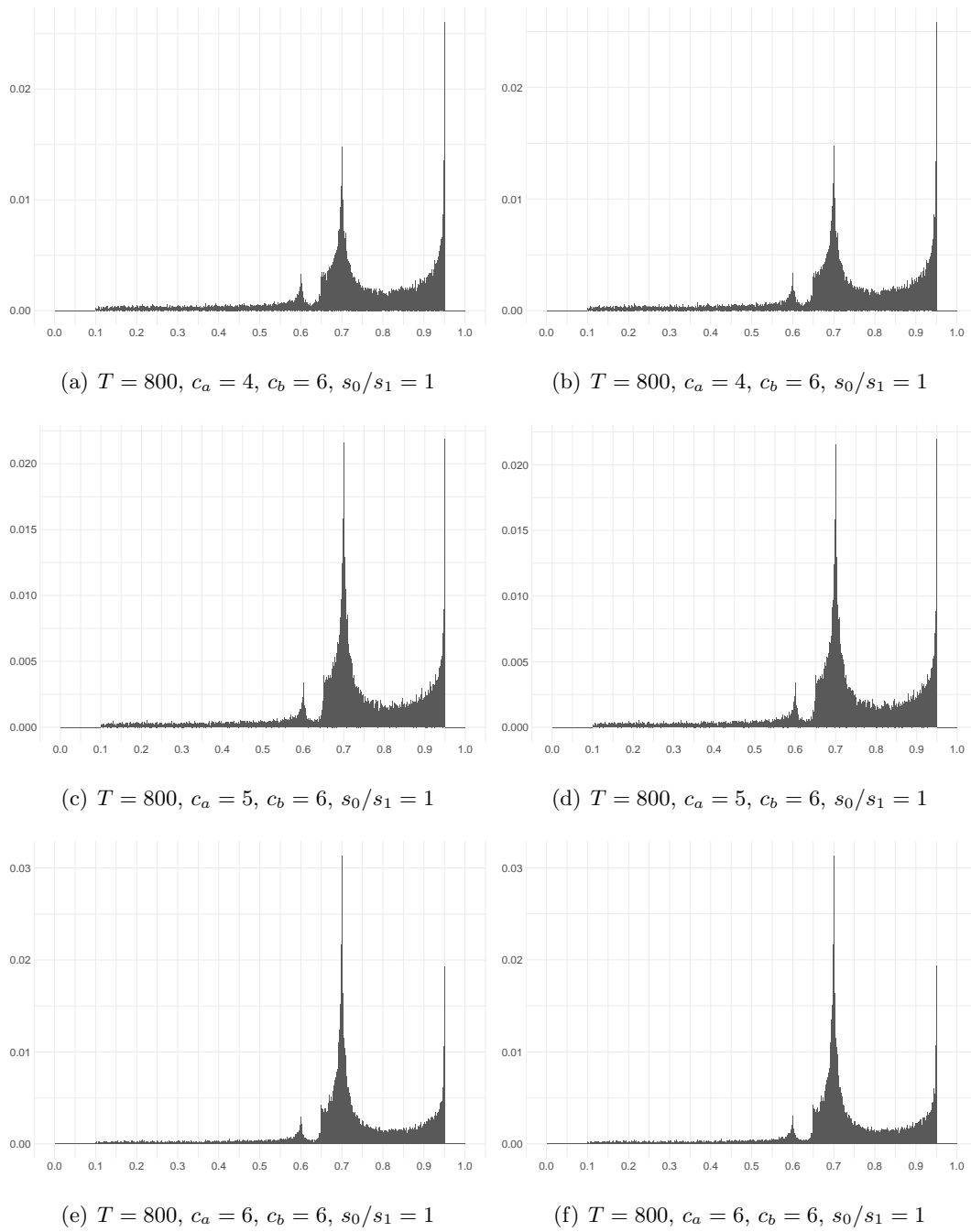


Figure C.4: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 1$ ,  $T = 800$

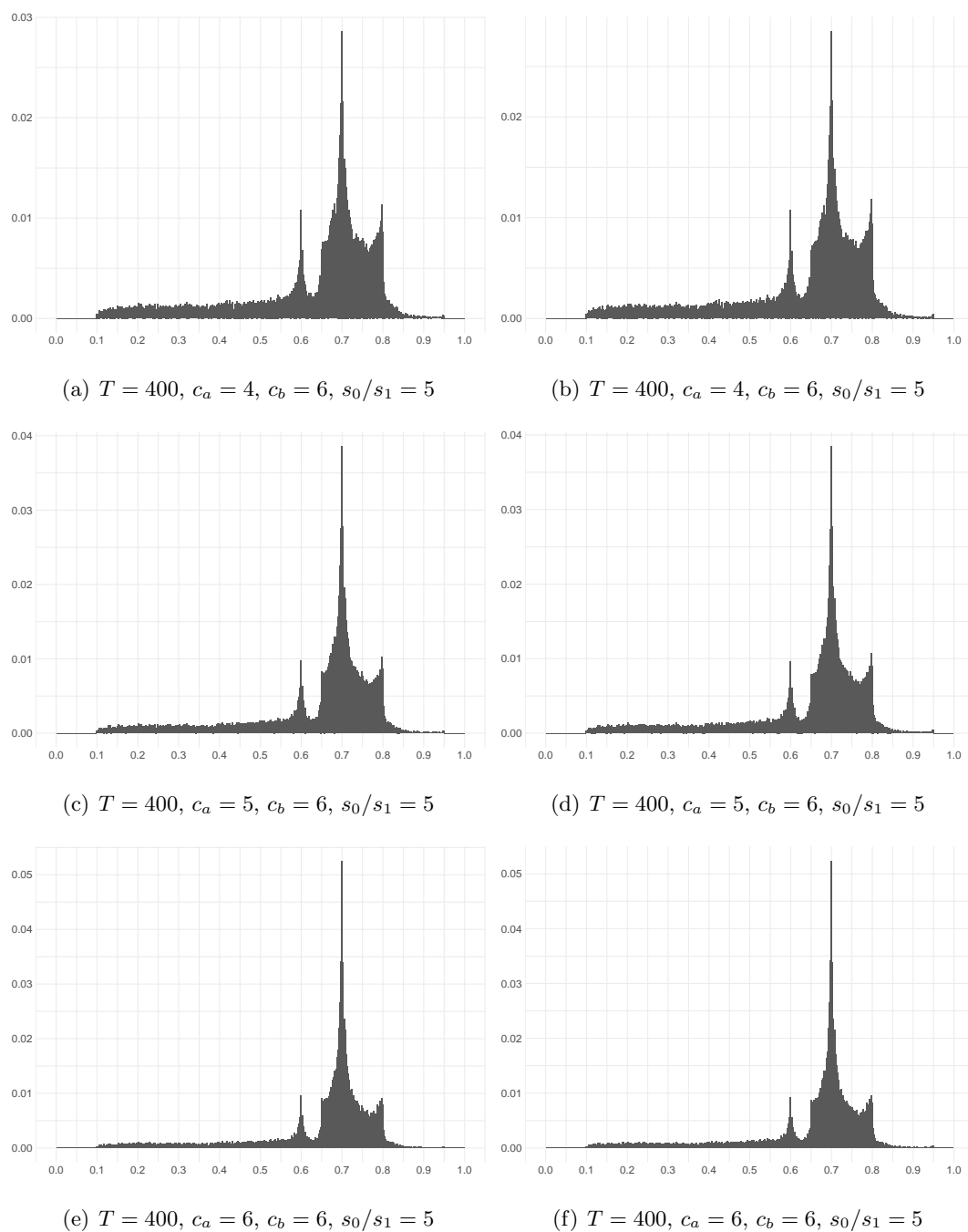


Figure C.5: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 5$ ,  $T = 400$

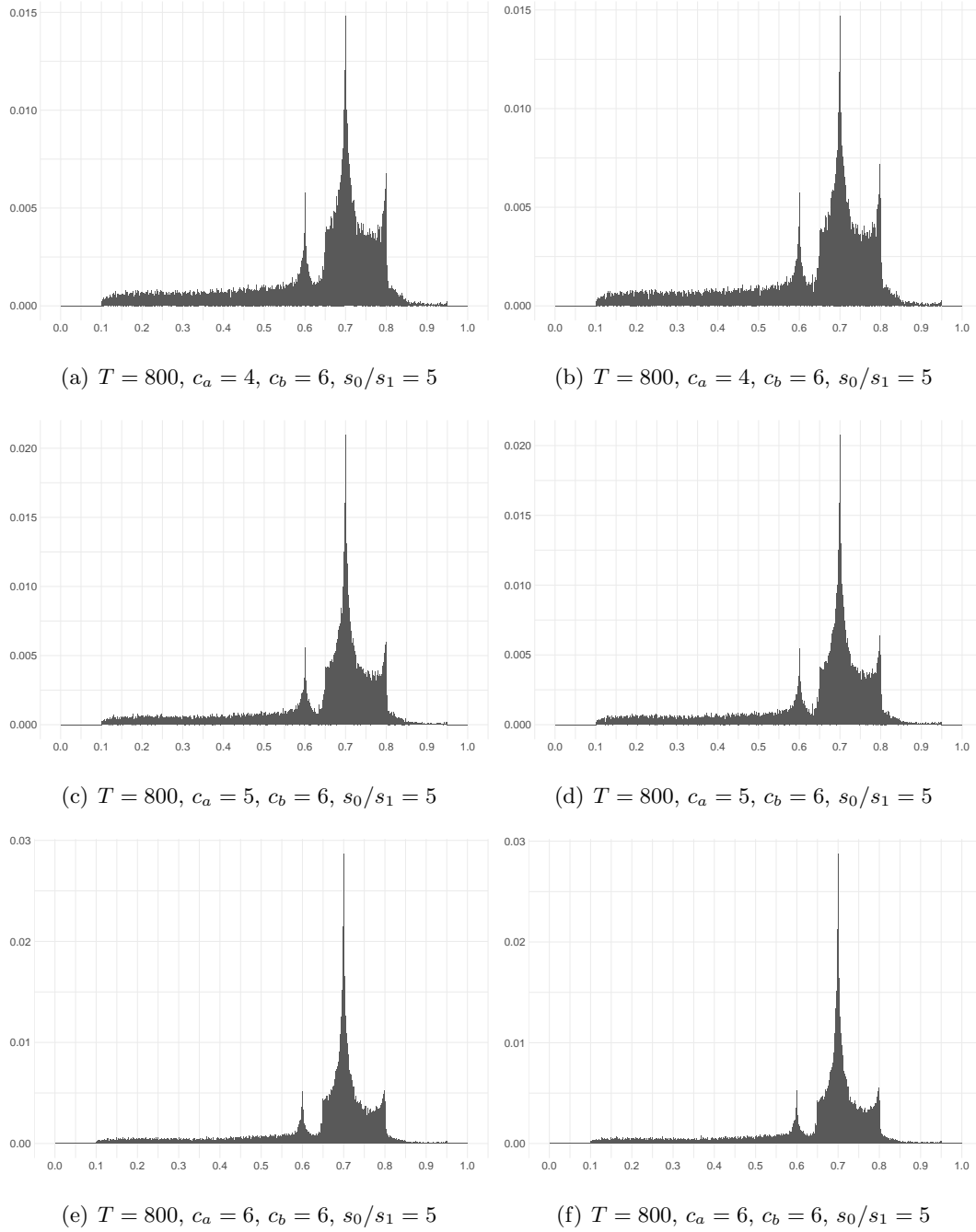


Figure C.6: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.8$ ,  $s_0/s_1 = 5$ ,  $T = 800$

D.  $\tau = 0.2, \hat{k}_c$

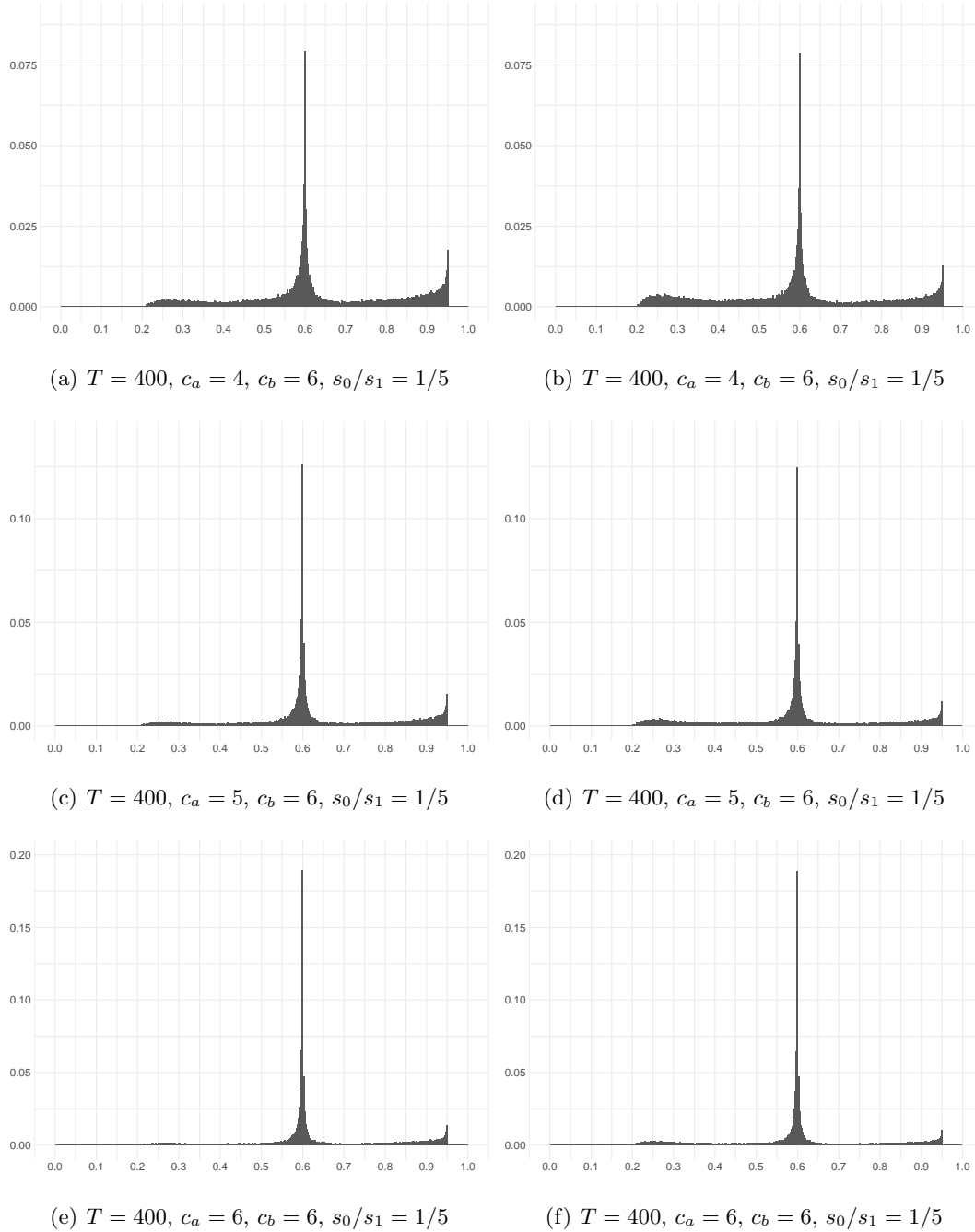


Figure D.1: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$



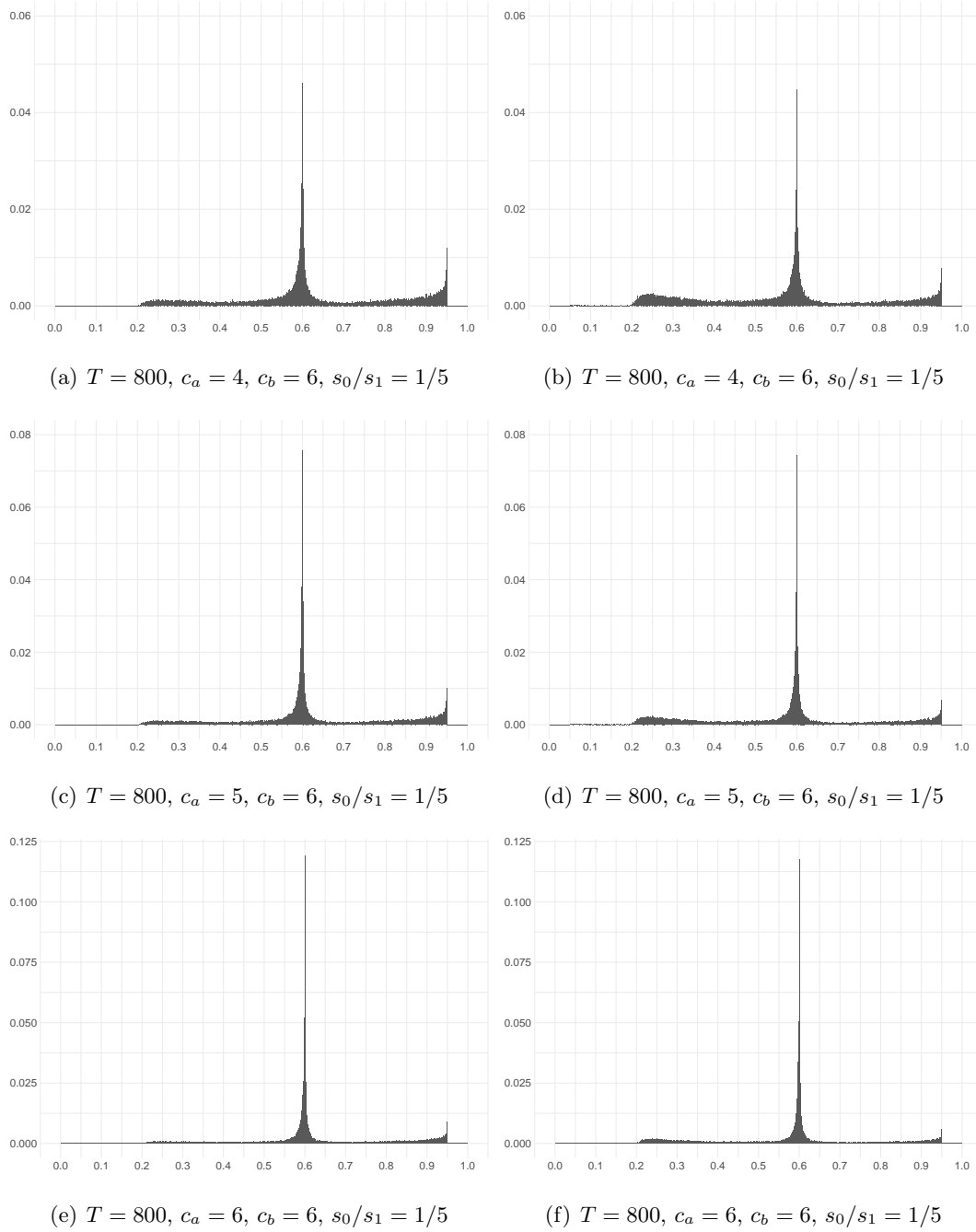


Figure D.2: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1/5$ ,  $T = 800$

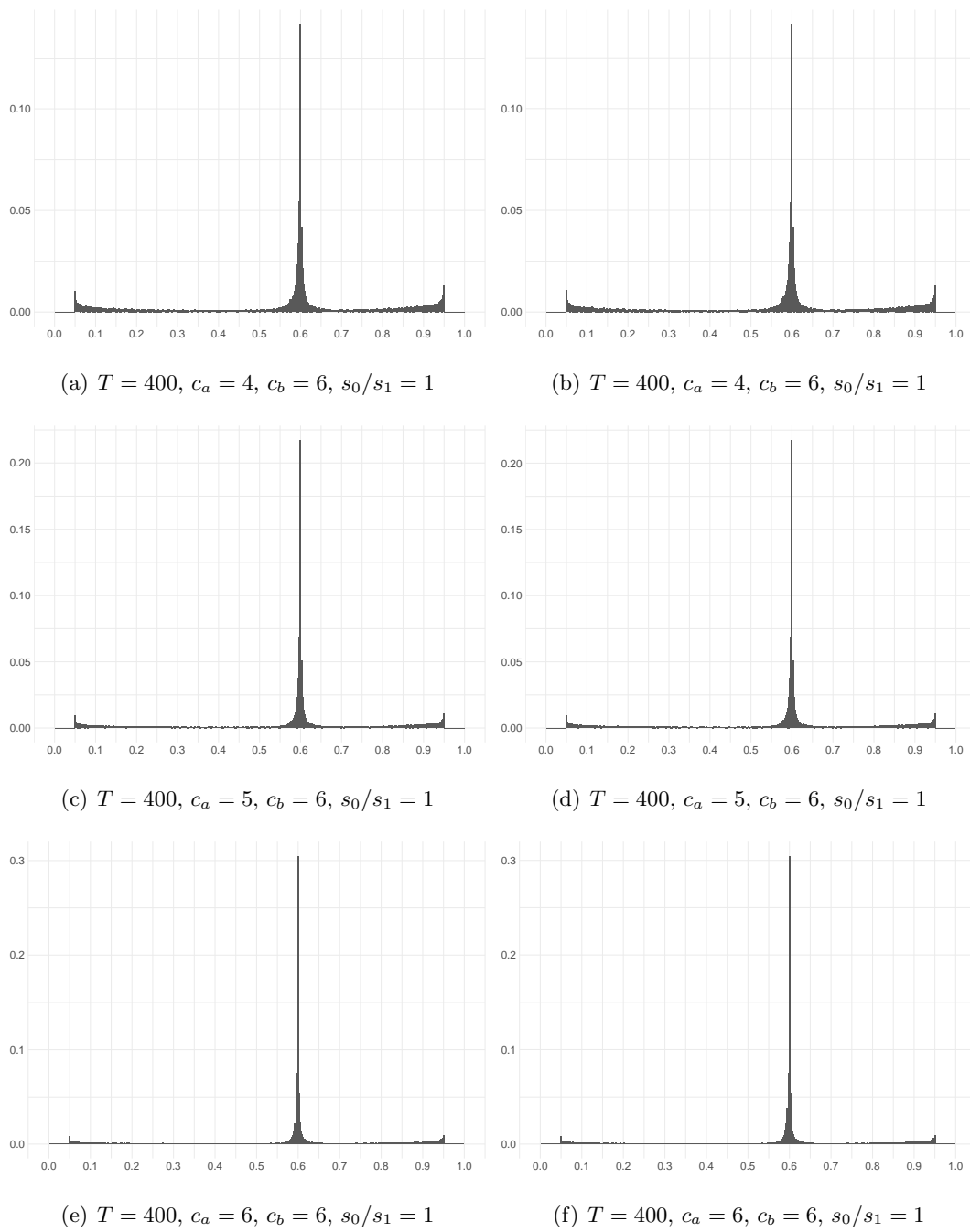


Figure D.3: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1$ ,  $T = 400$

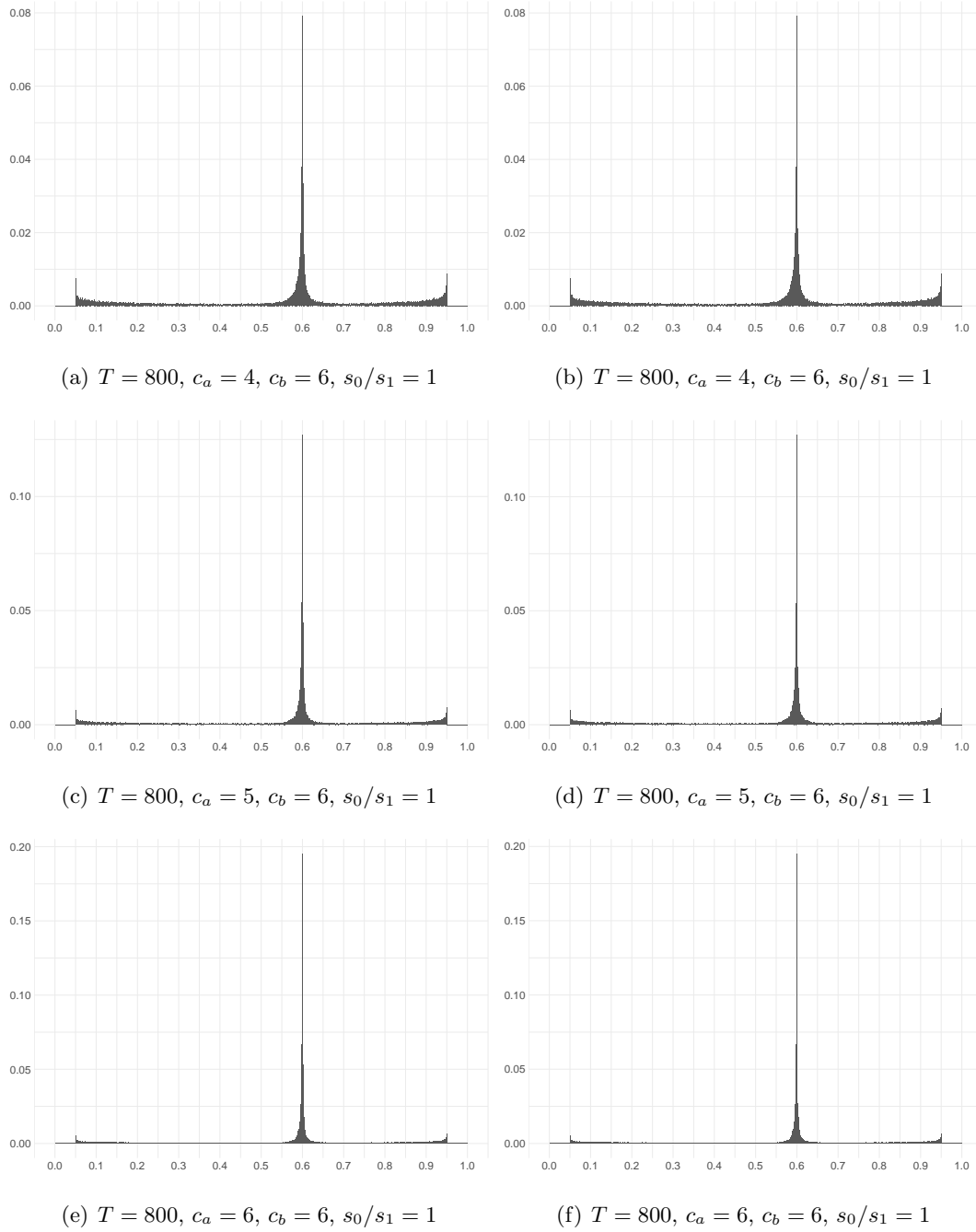


Figure D.4: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1$ ,  $T = 800$

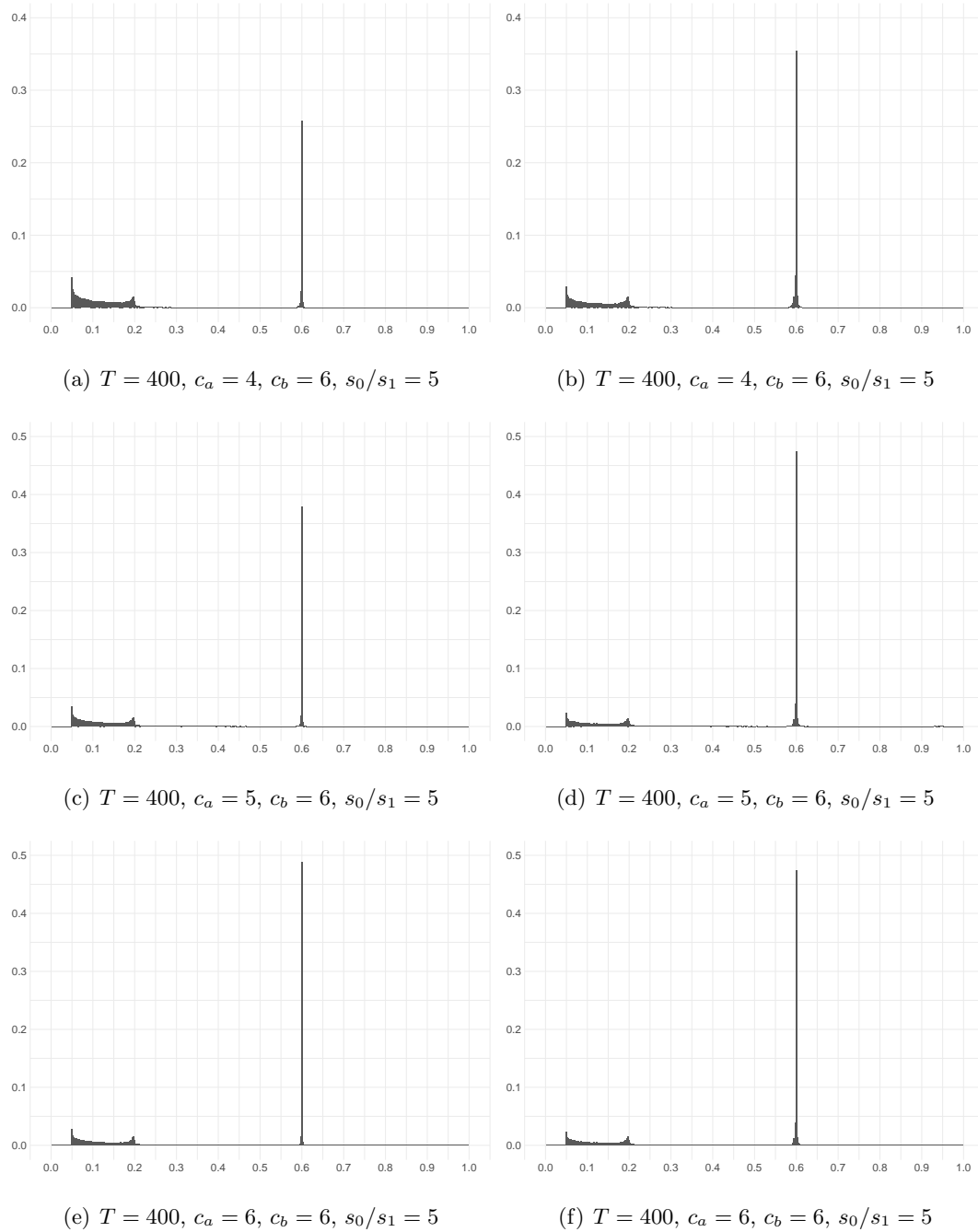


Figure D.5: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 400$

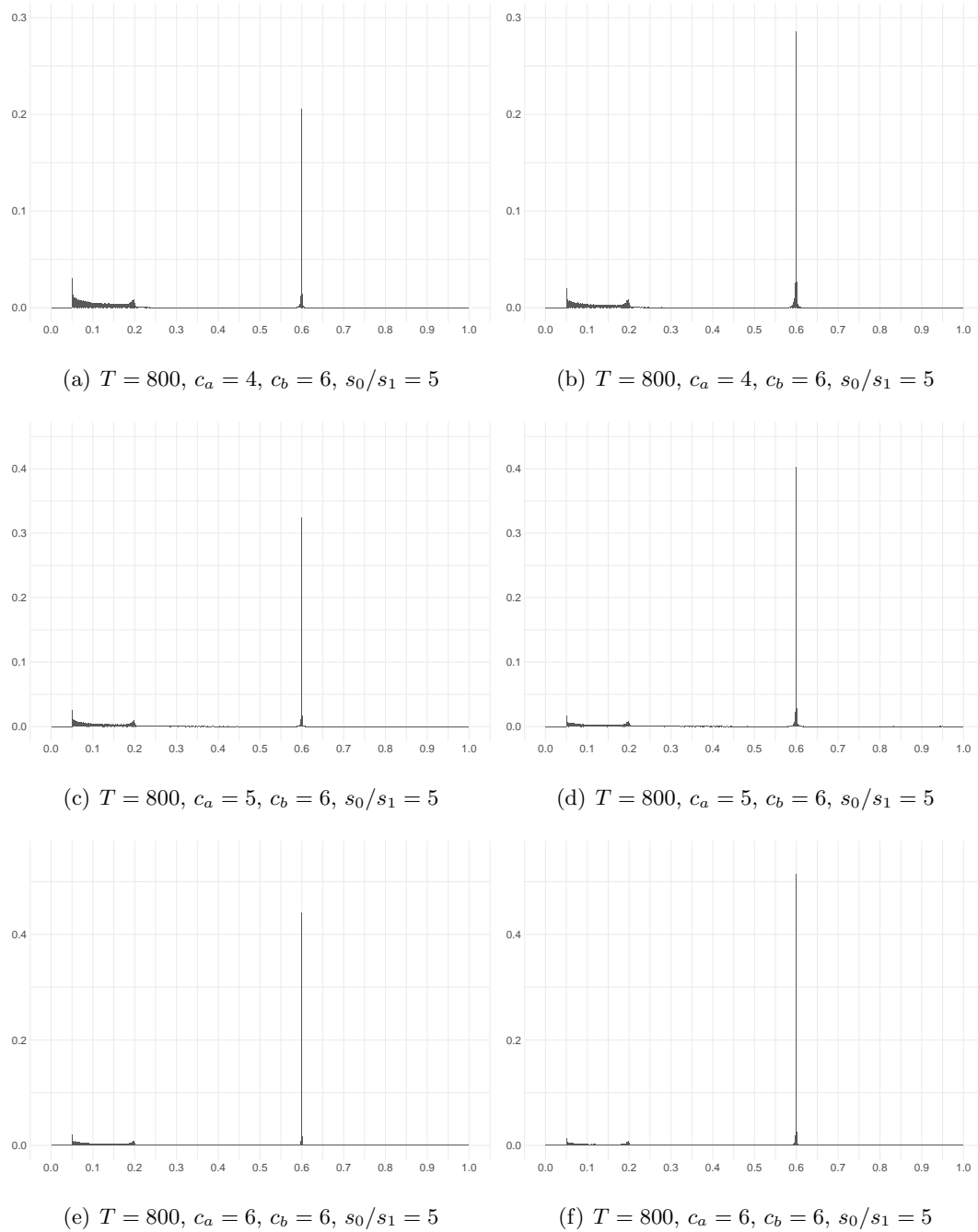
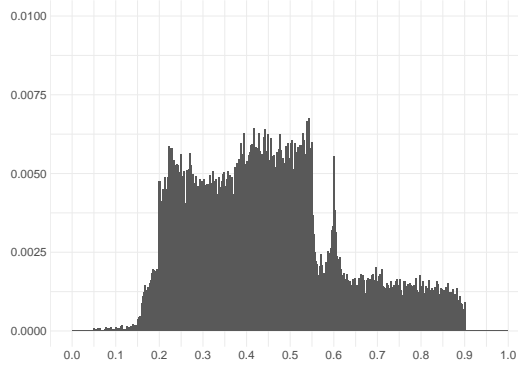
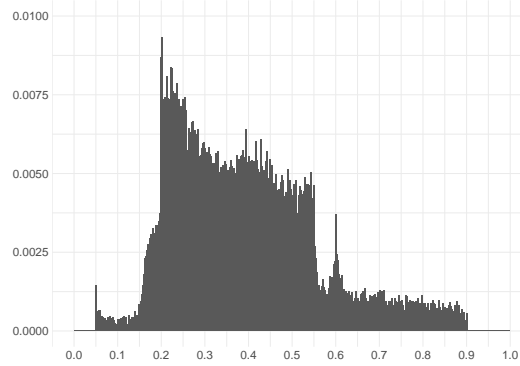


Figure D.6: Histograms of  $\hat{k}_c$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 800$

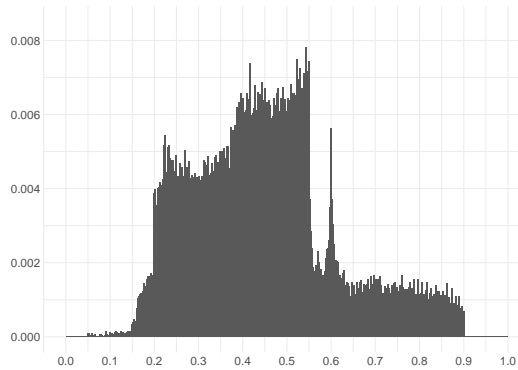
**E.**  $\tau = 0.2, \hat{k}_e$



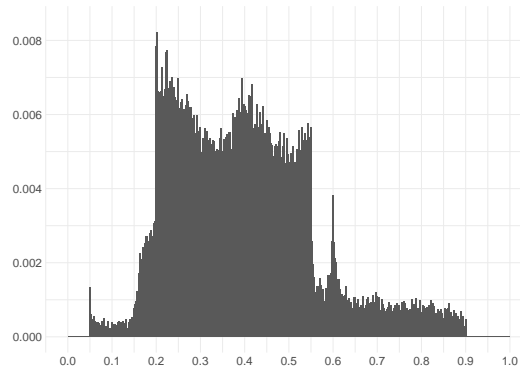
(a)  $T = 400, c_a = 4, c_b = 6, s_0/s_1 = 1/5$



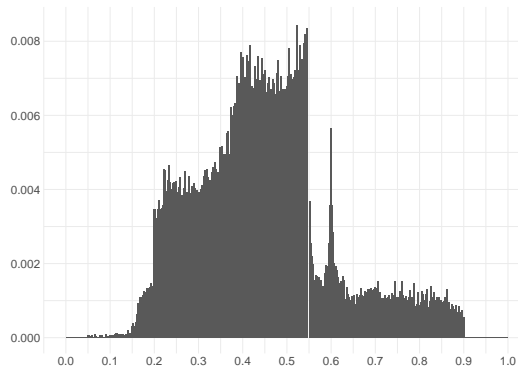
(b)  $T = 400, c_a = 4, c_b = 6, s_0/s_1 = 1/5$



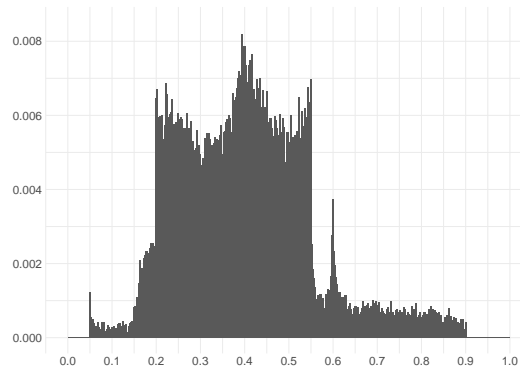
(c)  $T = 400, c_a = 5, c_b = 6, s_0/s_1 = 1/5$



(d)  $T = 400, c_a = 5, c_b = 6, s_0/s_1 = 1/5$



(e)  $T = 400, c_a = 6, c_b = 6, s_0/s_1 = 1/5$



(f)  $T = 400, c_a = 6, c_b = 6, s_0/s_1 = 1/5$

Figure E.1: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$

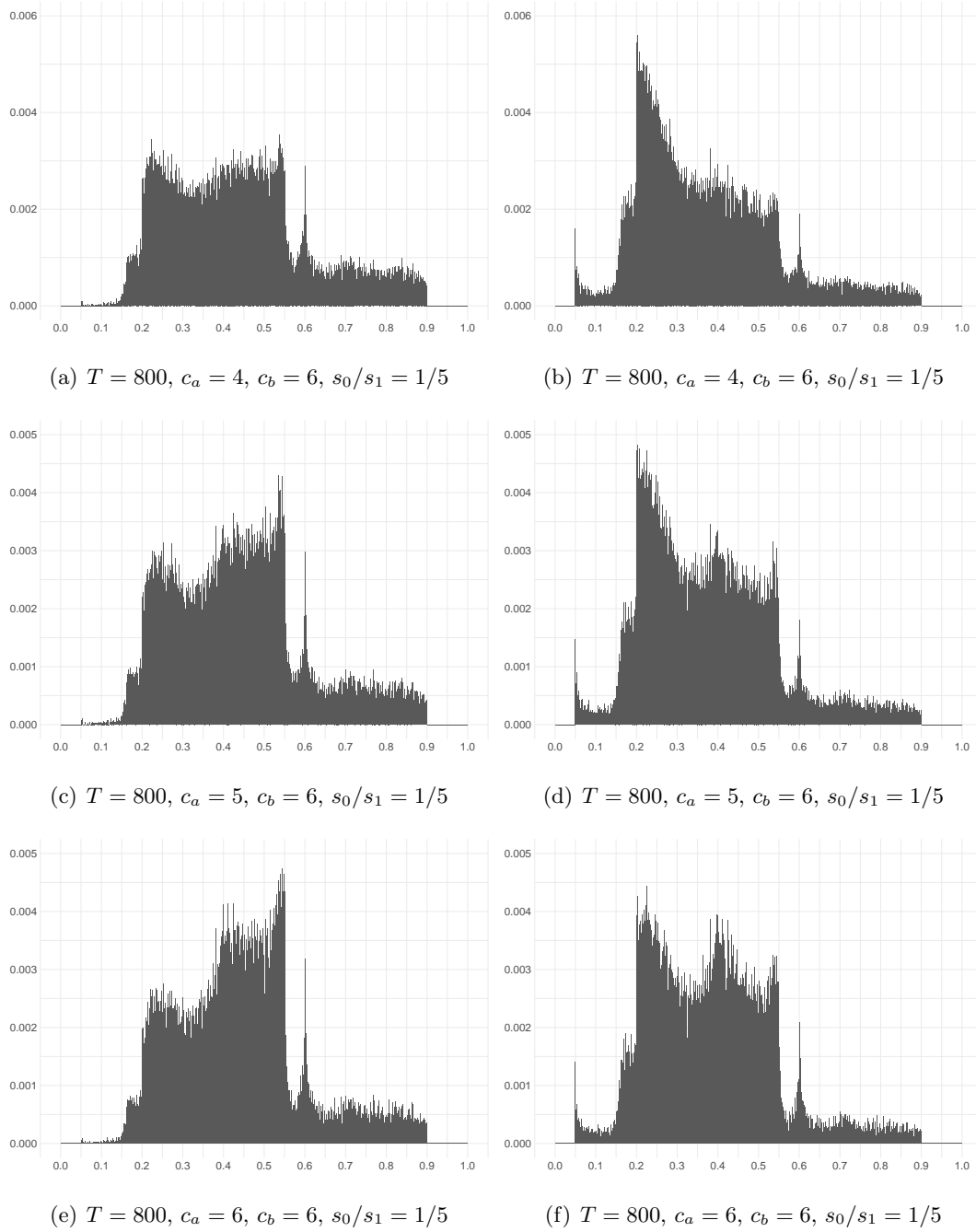


Figure E.2: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1/5$ ,  $T = 800$

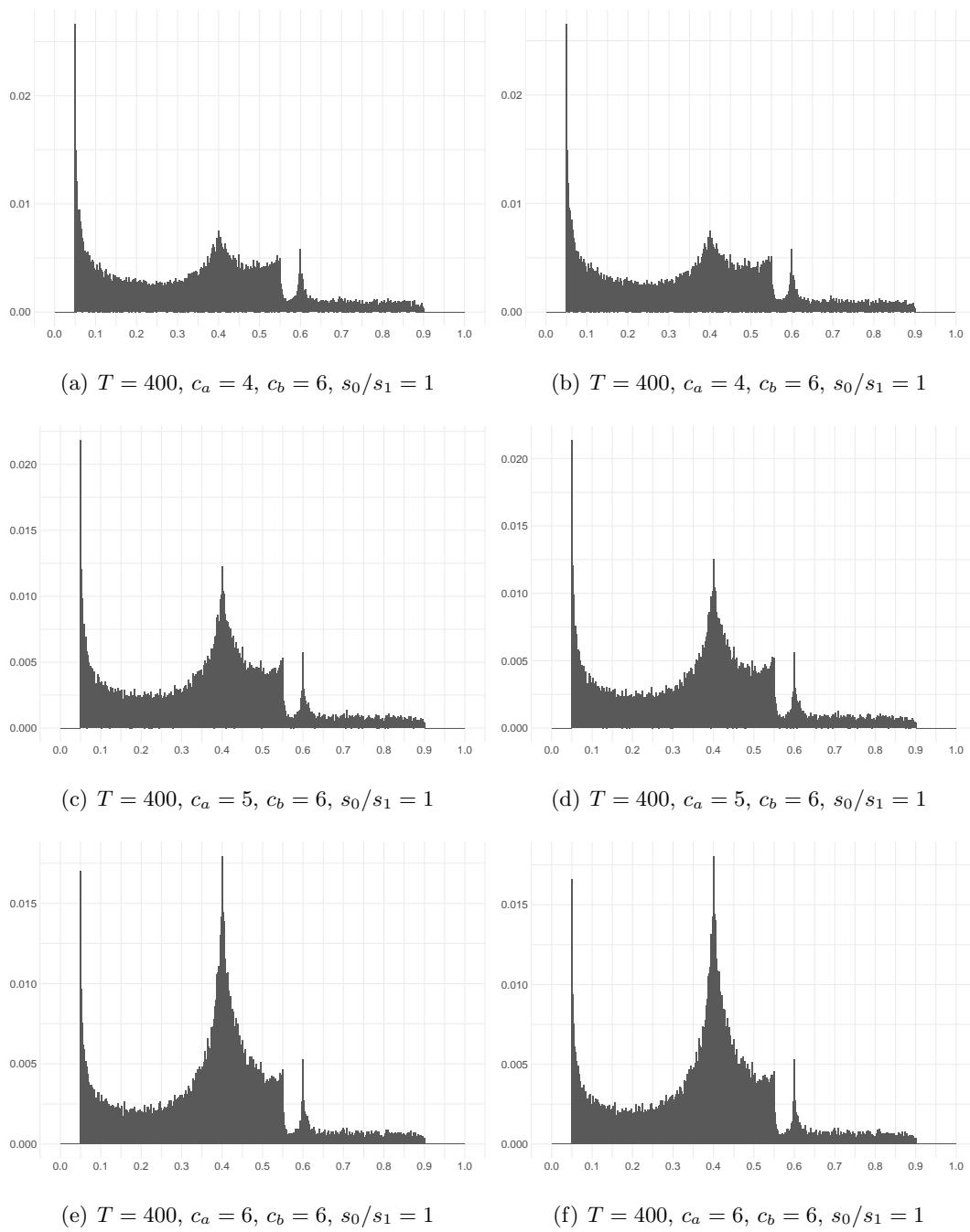


Figure E.3: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1$ ,  $T = 400$



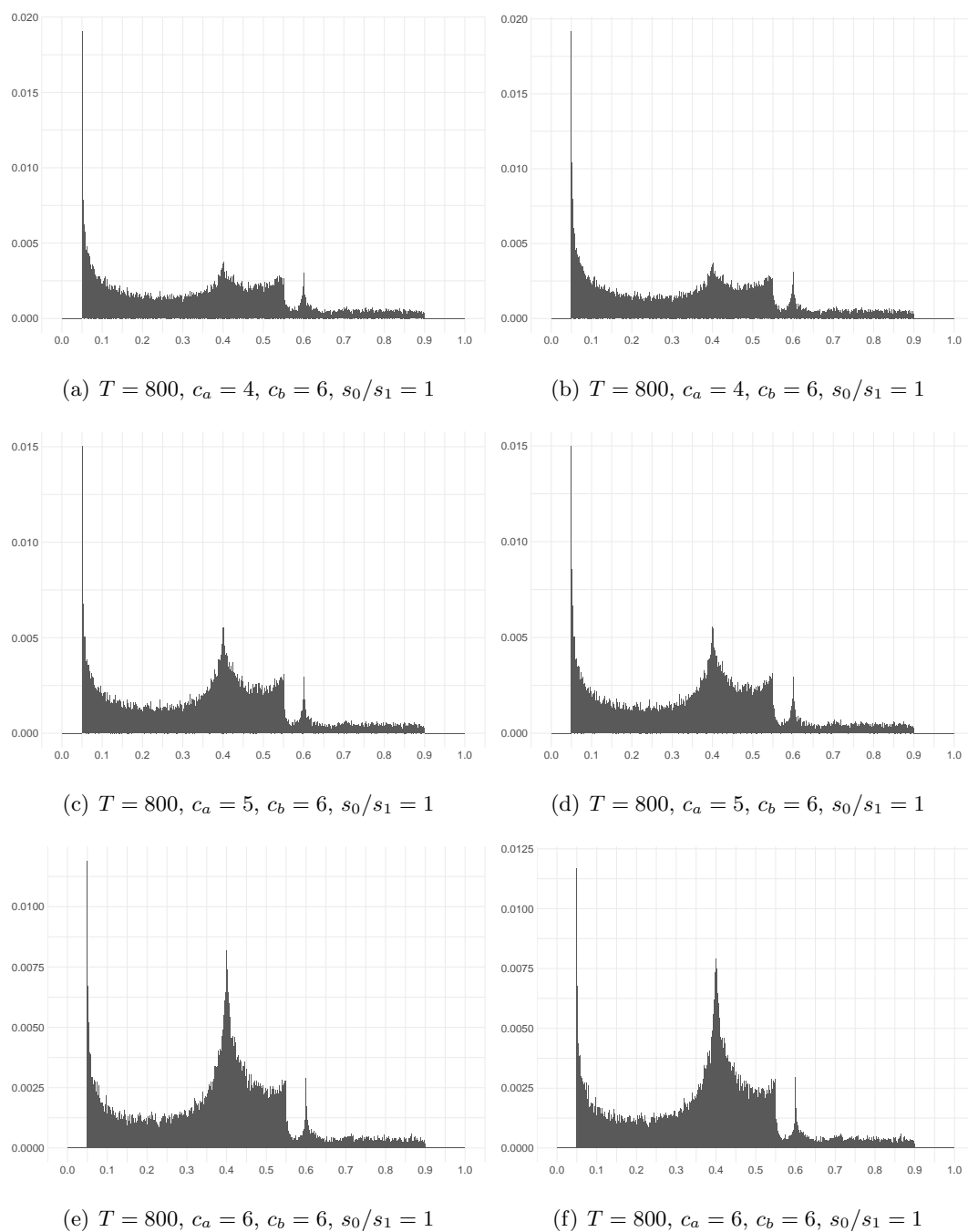


Figure E.4: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1$ ,  $T = 800$

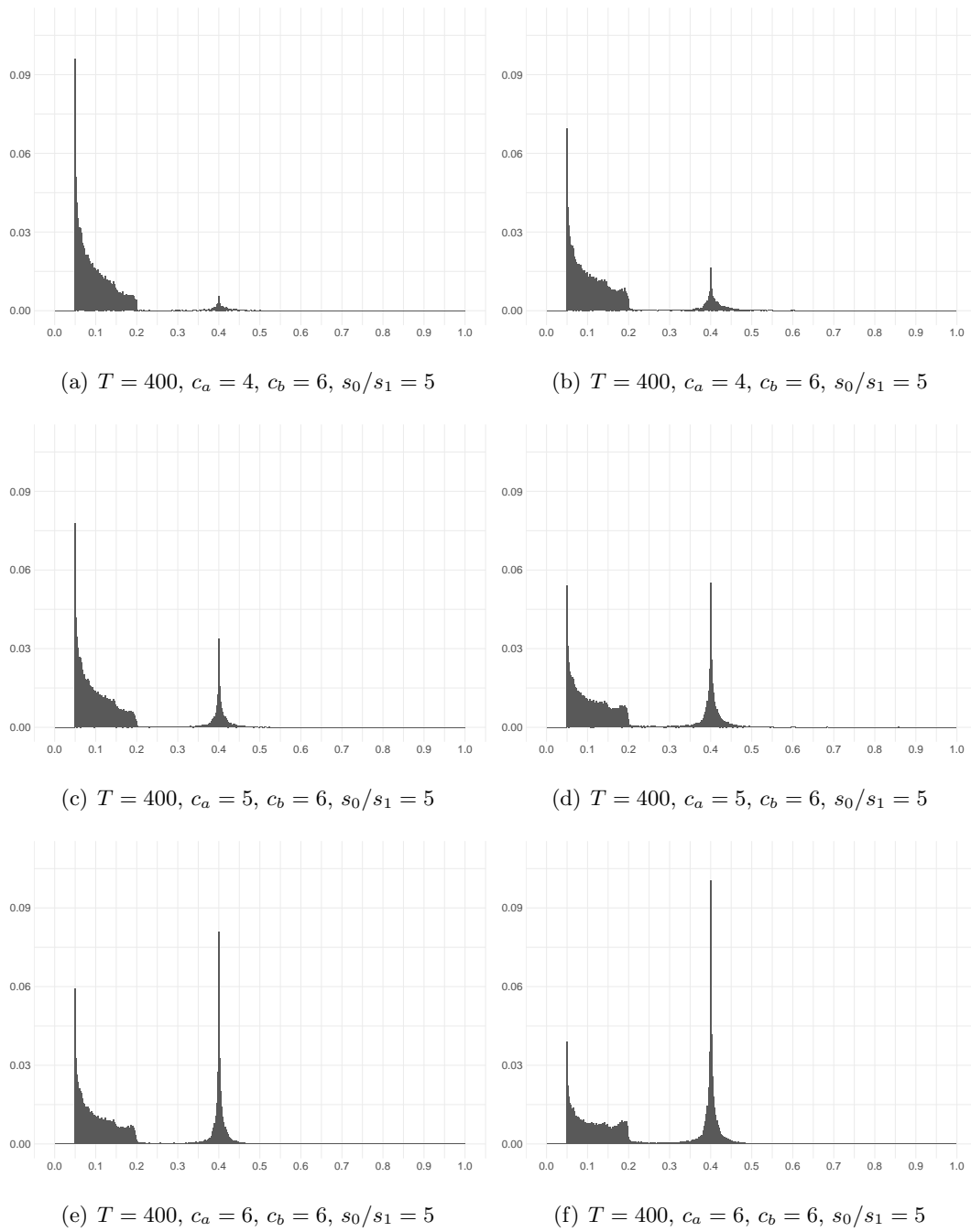


Figure E.5: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 400$

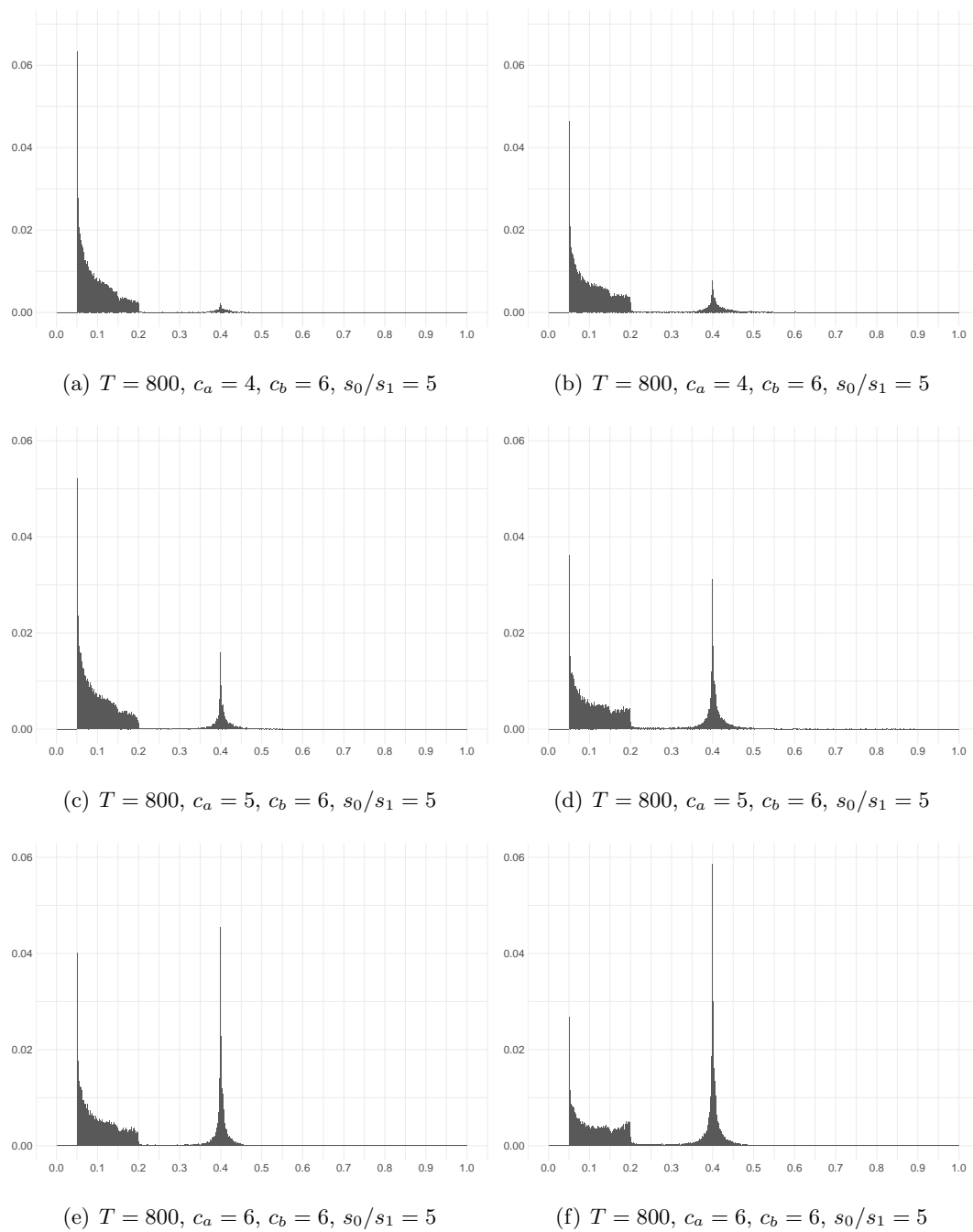


Figure E.6: Histograms of  $\hat{k}_e$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 800$

**F.**  $\tau = 0.2, \hat{k}_r$

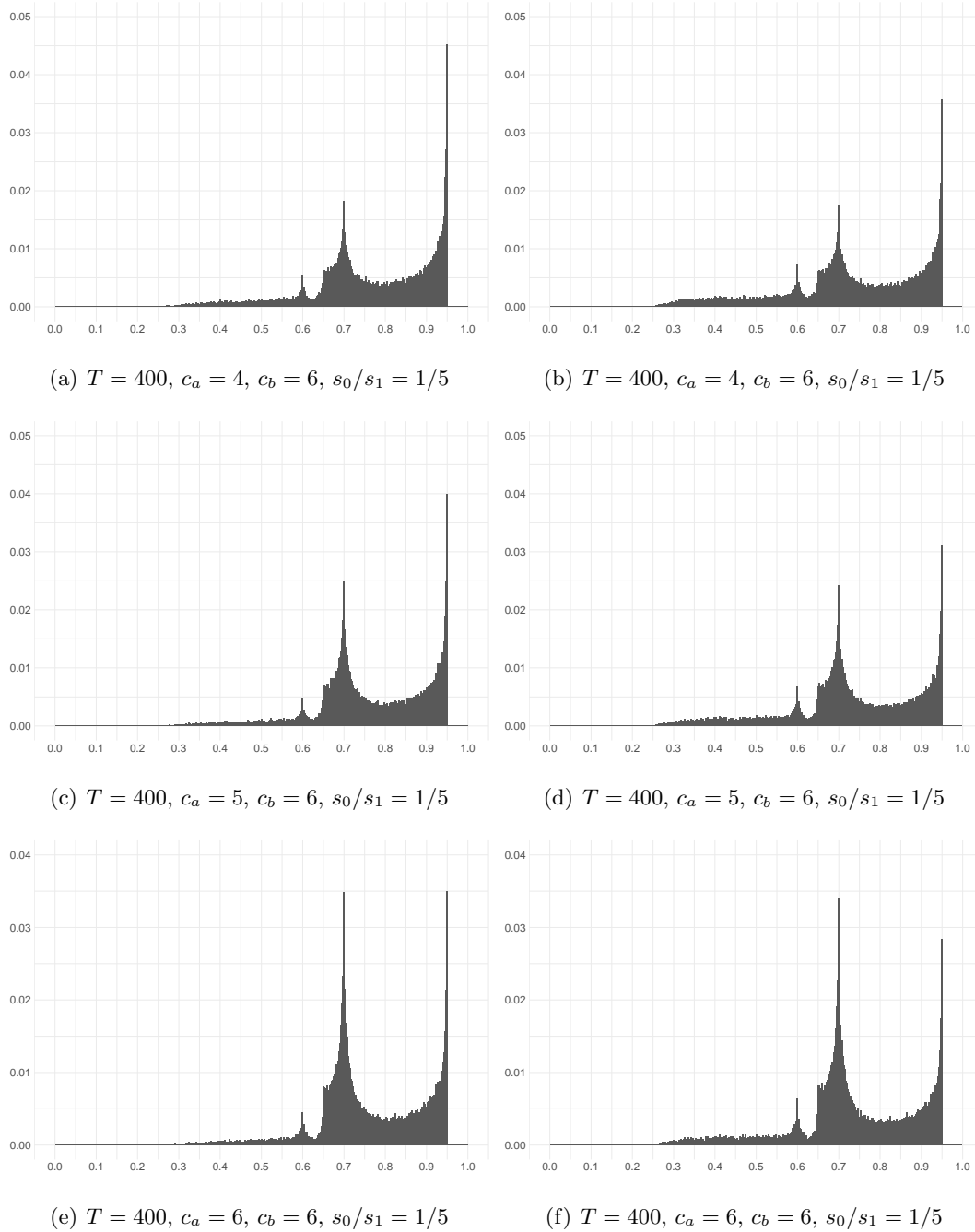


Figure F.1: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1/5$ ,  $T = 400$

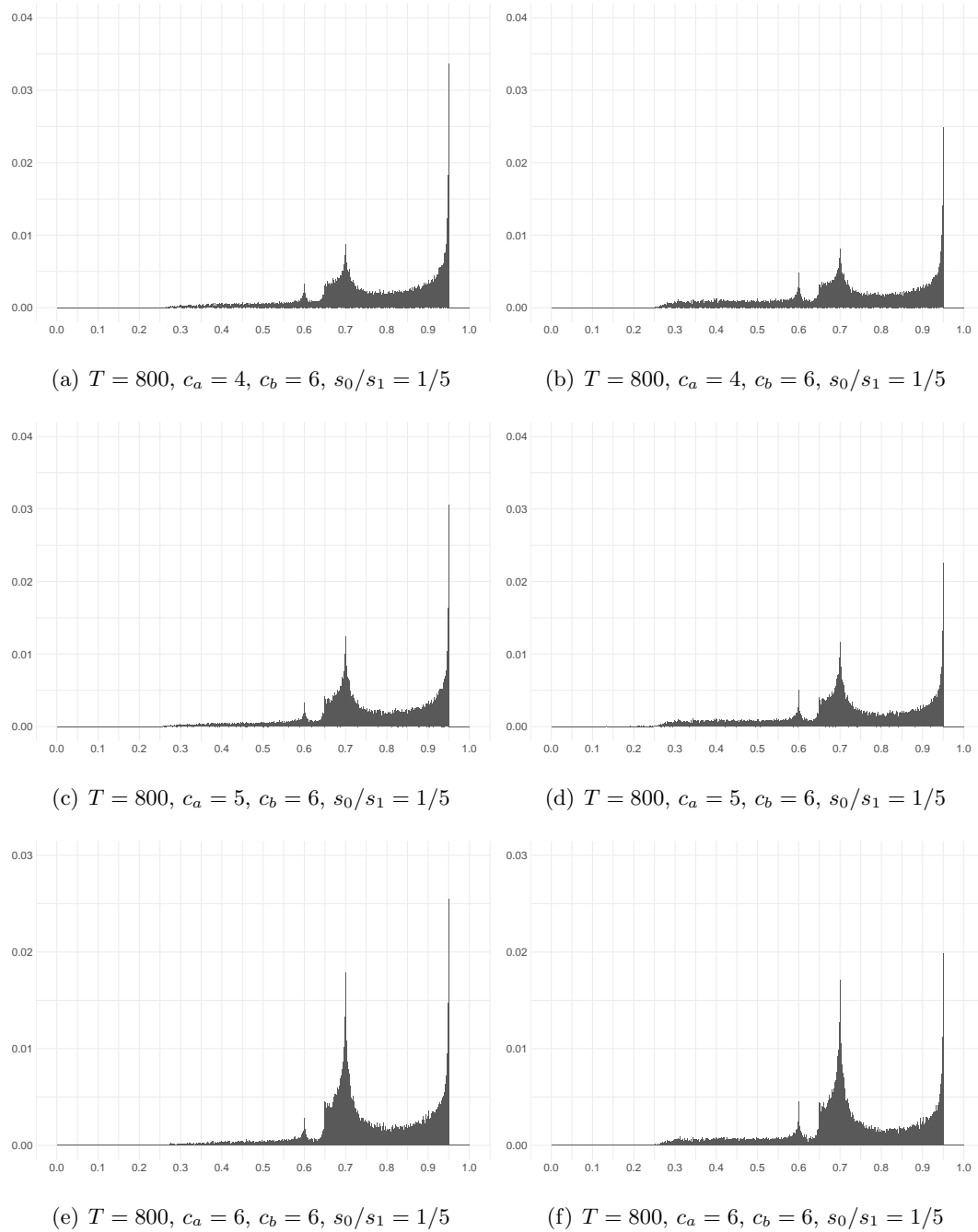


Figure F.2: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1/5$ ,  $T = 800$

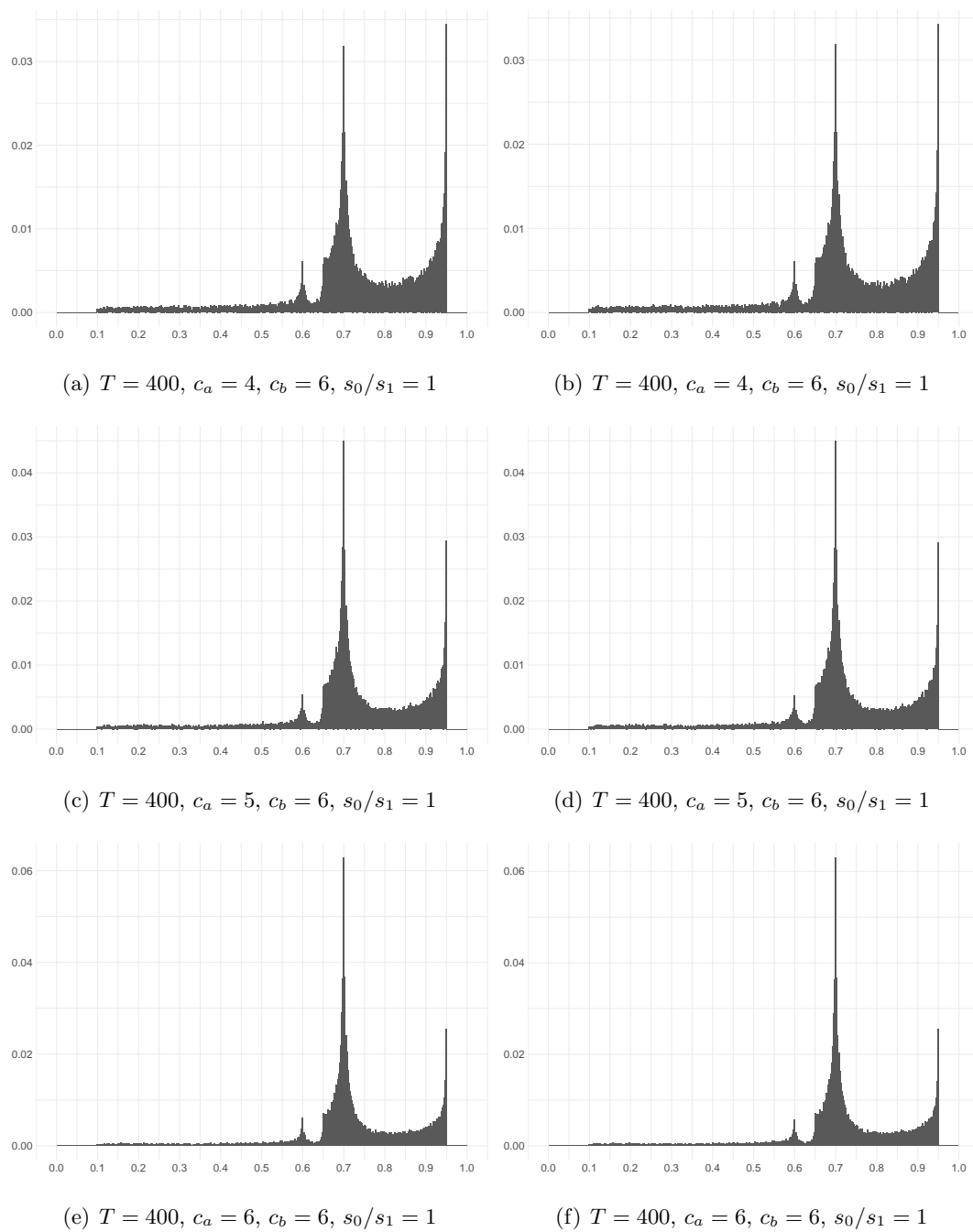


Figure F.3: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1$ ,  $T = 400$

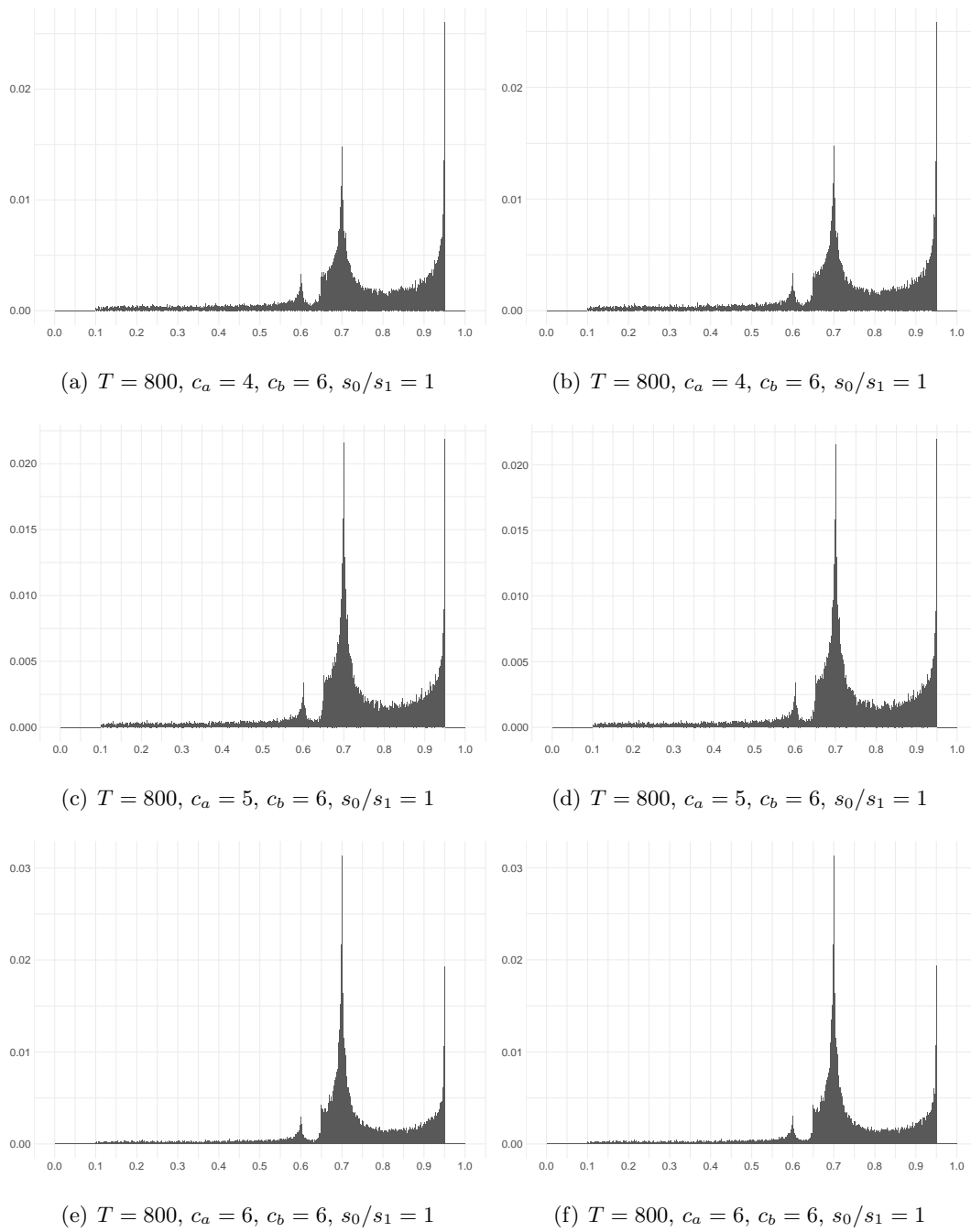


Figure F.4: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 1$ ,  $T = 800$

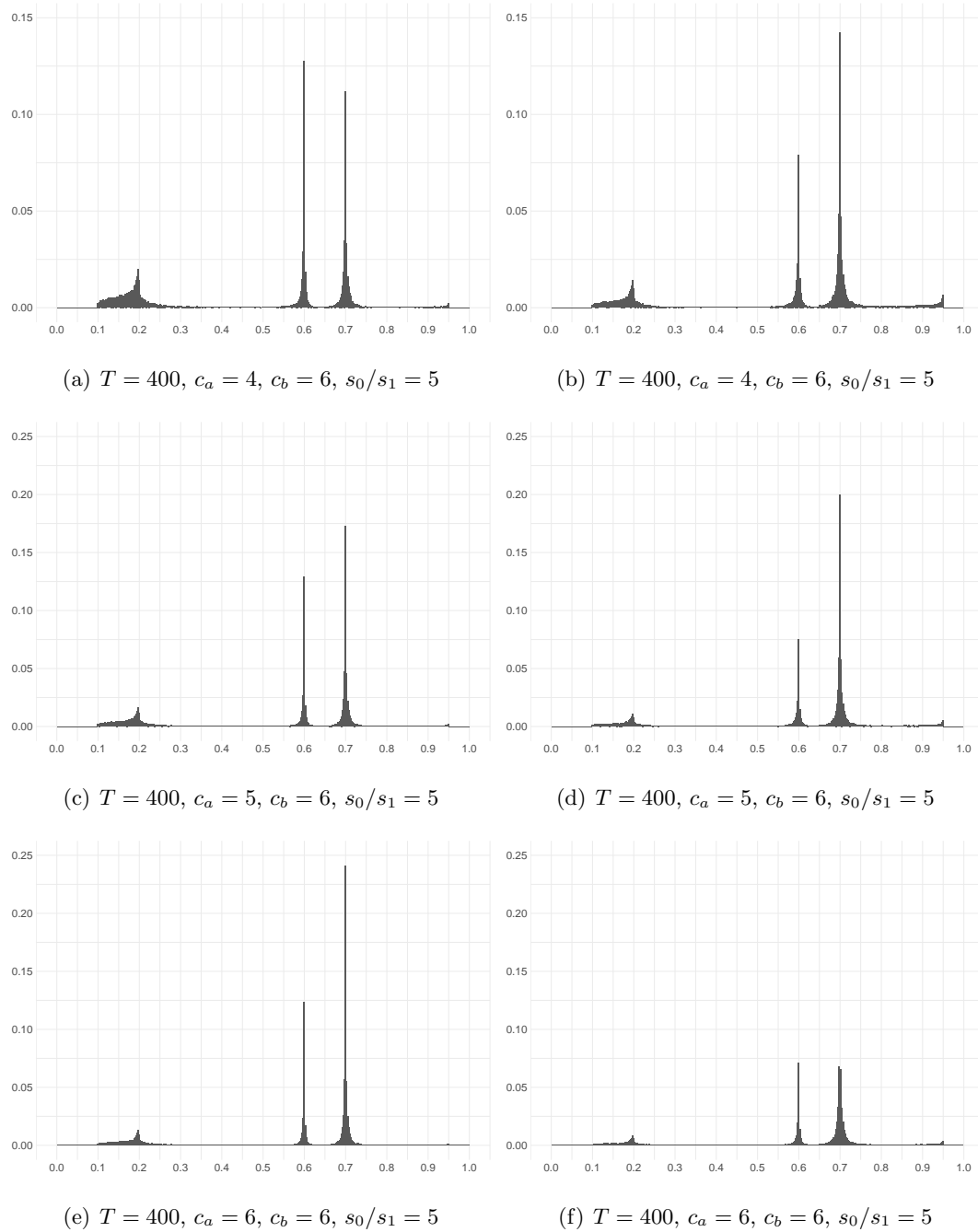


Figure F.5: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 400$



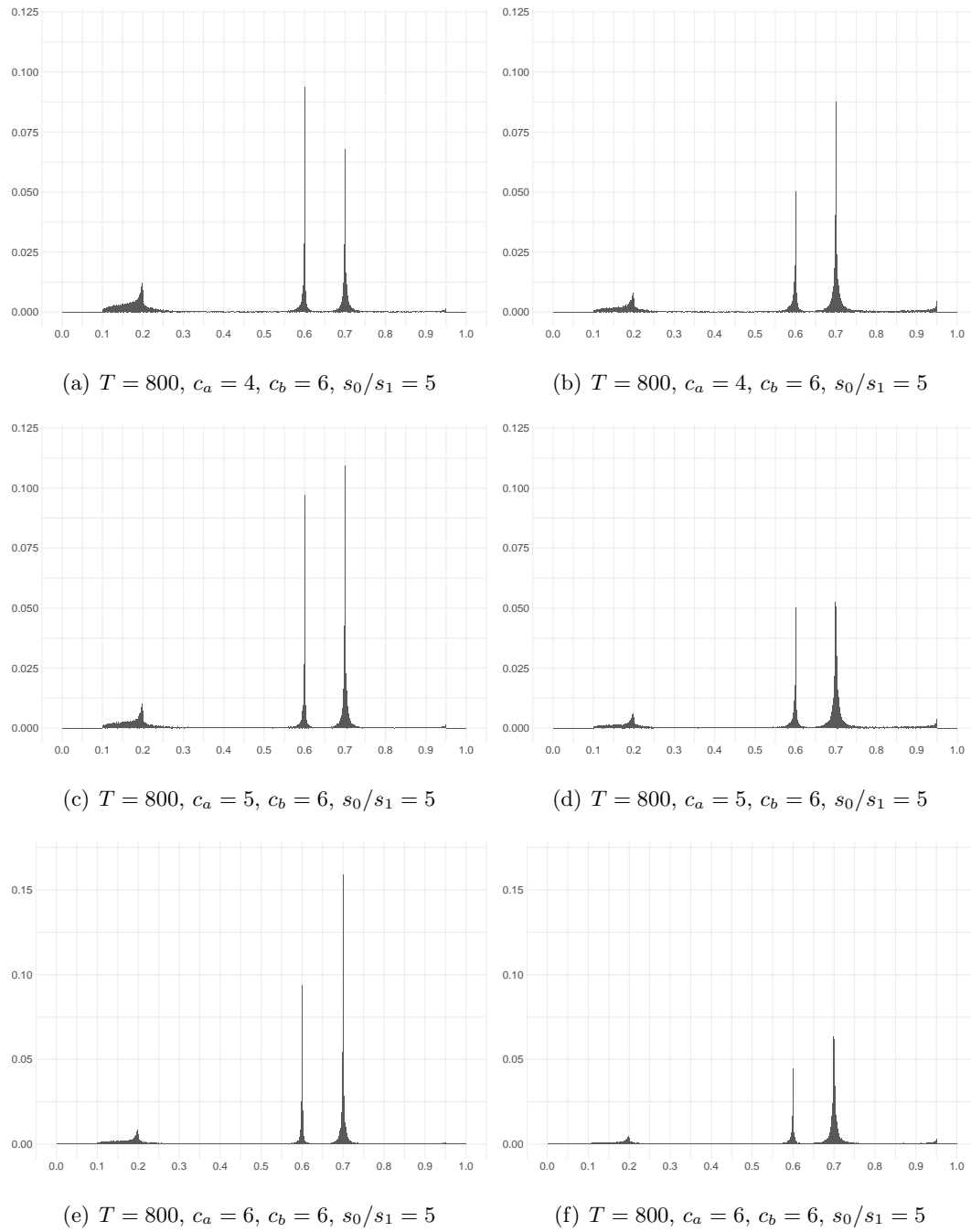


Figure F.6: Histograms of  $\hat{k}_r$  for  $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$ ,  $\tau = 0.2$ ,  $s_0/s_1 = 5$ ,  $T = 800$