Appendix
Exact Mean Integrated Squared Error
of Higher-Order Kernel Estimators:

Bruce E. Hansen
University of Wisconsin

June 2003

1 Numerical Integration

The integrals $I_1(h, K, f)$ and $I_2(h, K, f)$ were calculated both numerically and by the series expansions from Theorems 5 and 6 for the Marron-Wand densities. The Gaussian kernels G_{2r}, the polynomial kernels $K_{2r,s}$, the Dirichlet D and the trapezoid λ_2 kernels were evaluated, the first two for a range of values of r and s.

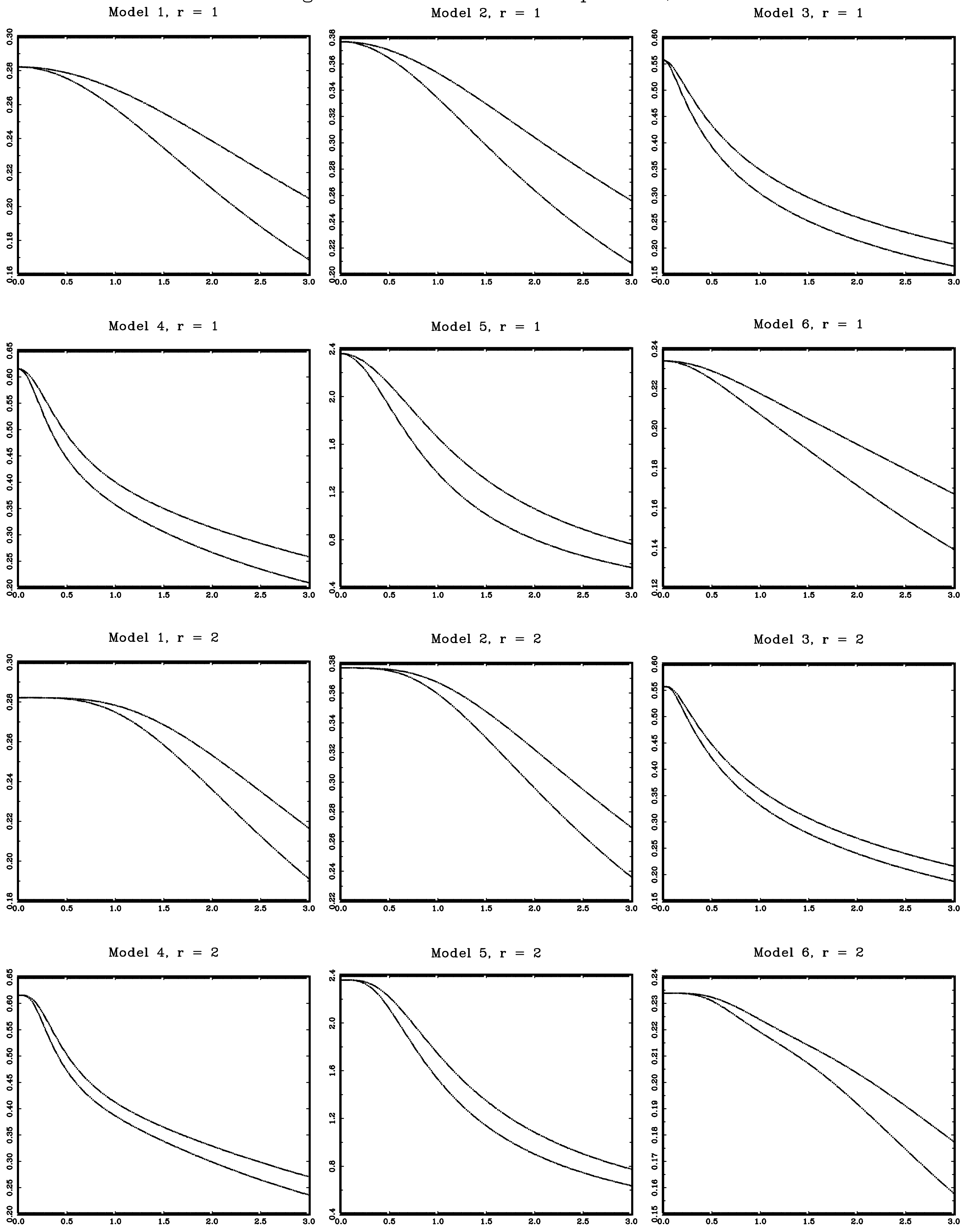
The following seven pages display these calculations. For simplicity we only display results for the first six models, for r from 1 to 6, and for $s = 2$. Each figure displays the integrals $I_1(h, K, f)$ and $I_2(h, K, f)$ for a specific model and kernel, displayed as functions of h on the x-axis. The range for h is the same as for the calculation of MISE in the next section. The numerical calculation is displayed using a dotted line, and the series expansion calculation using a dashed line. If the latter failed to converge for a value of h, the dashed line is missing. Typically the upper line is the integral $I_1(h, K, f)$ and the lower line is $I_2(h, K, f)$, but a few cases the lines intersect.

The first three pages are for the Gaussian kernels G_{2r}. In all cases the numerical integrals and series expansions are virtually identical, so the dotted and dashed lines are coincident.

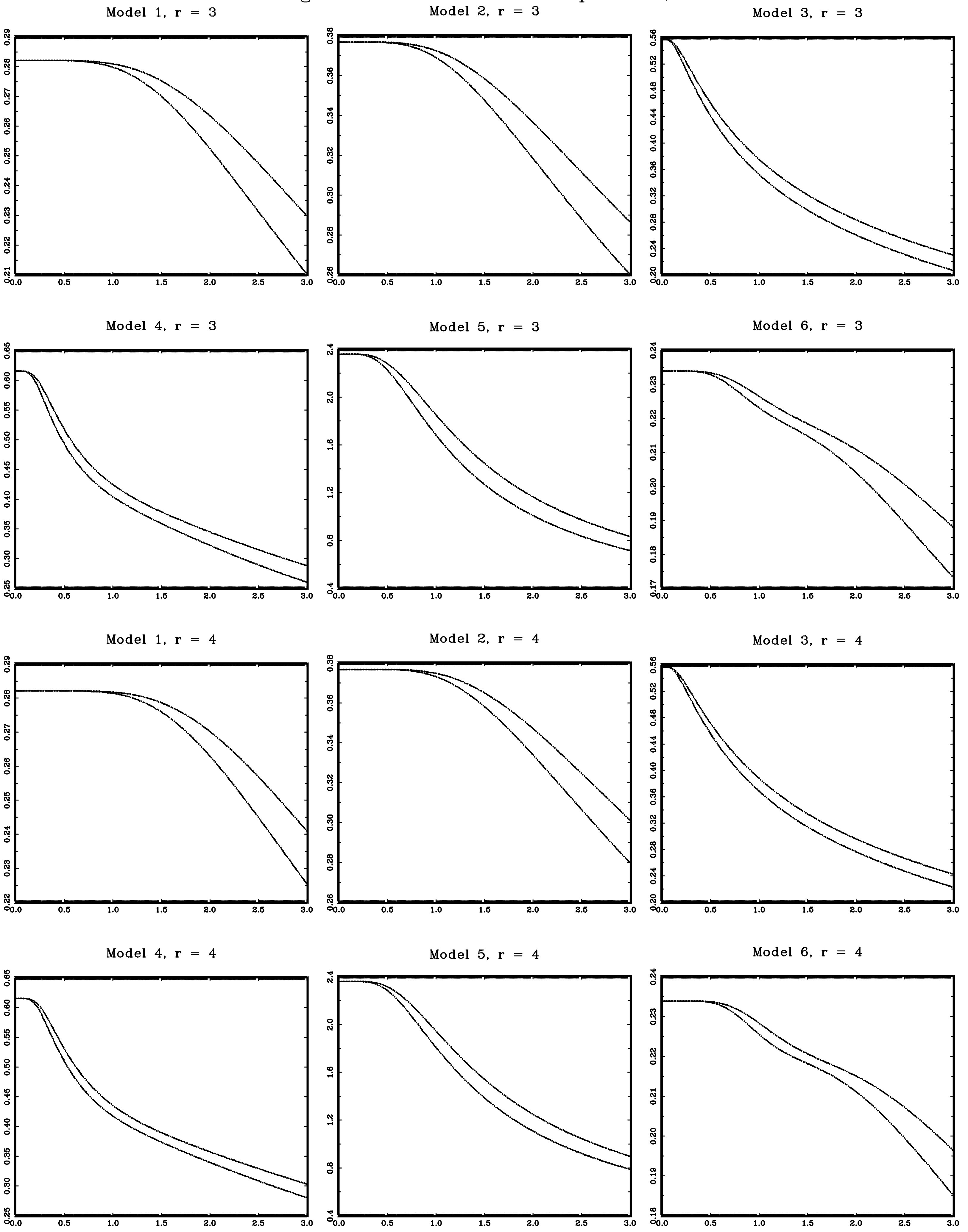
The following three pages are for the Polynomial kernels $K_{2r,s}$. In some cases (such as the first upper-left panel) the two calculations coincide for all bandwidths h. In other cases (such as the first upper-right panel) the two coincide for small values of h, and then the series expansion fails to converge for larger values. In other cases, the series expansion calculation "goes wild" for some values of h just smaller than values for which convergence fails. In both cases, the terms in the series expansions are explosive, with rounding error accumulating so that the sum is numerically unreliable.

The final (seventh) page displays the calculations for the Dirichlet and Trapezoid kernels. In all panels displayed, the two calculations coincide. There were models (not shown) where the series expansion failed to converge for very small bandwidths.

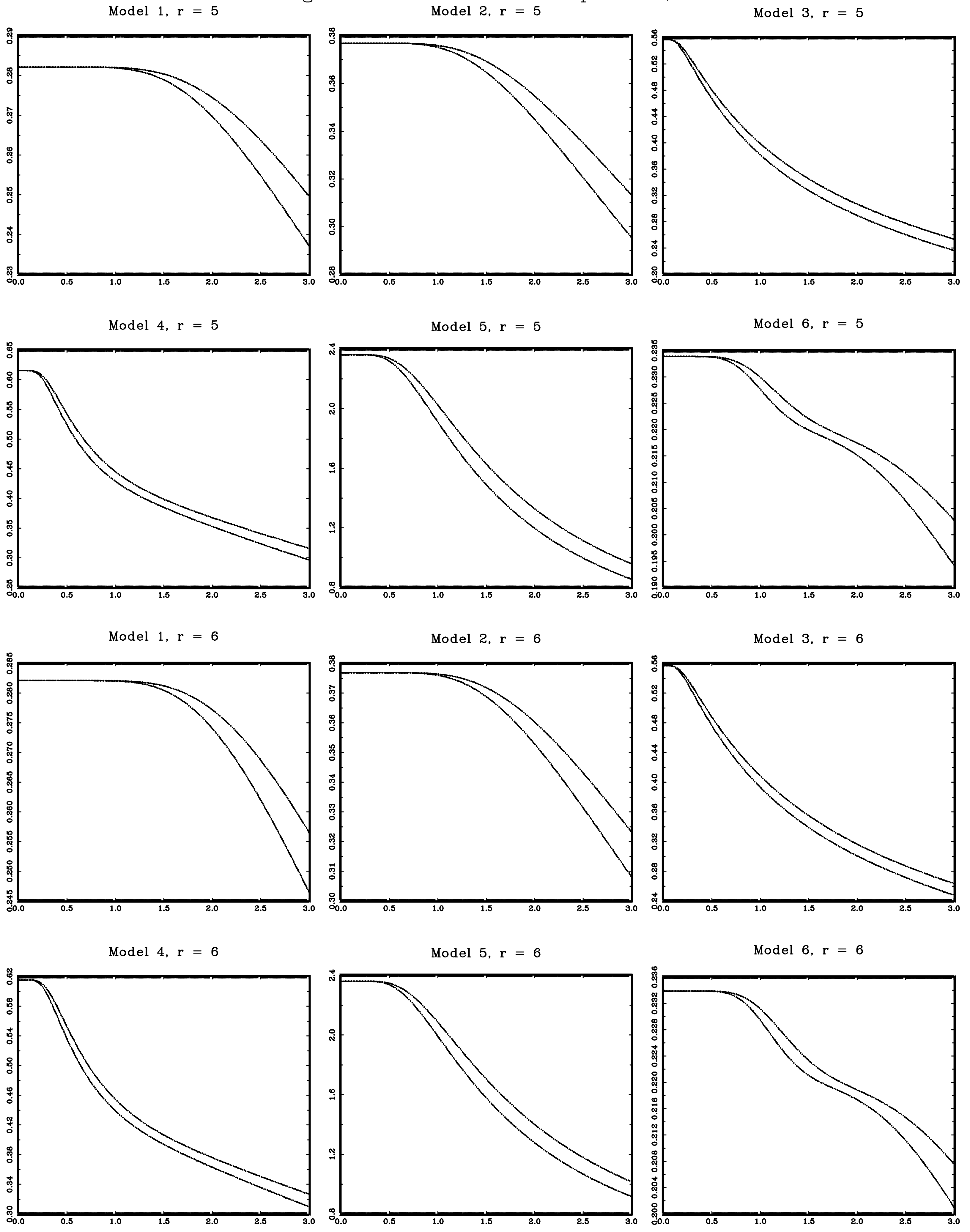
Numerical Integration Versus Series Expansion, Gaussian Kernel



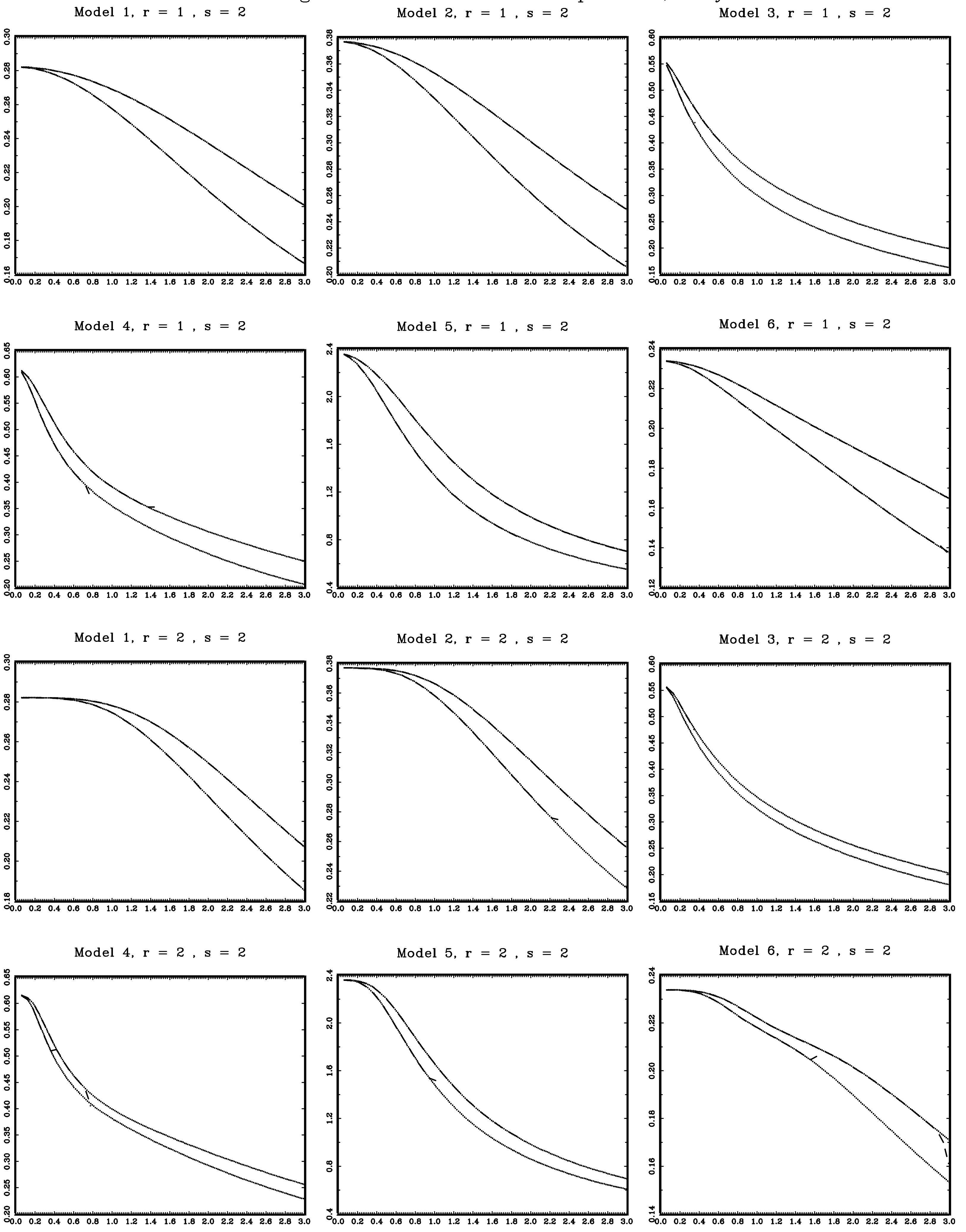
Numerical Integration Versus Series Expansion, Gaussian Kernel



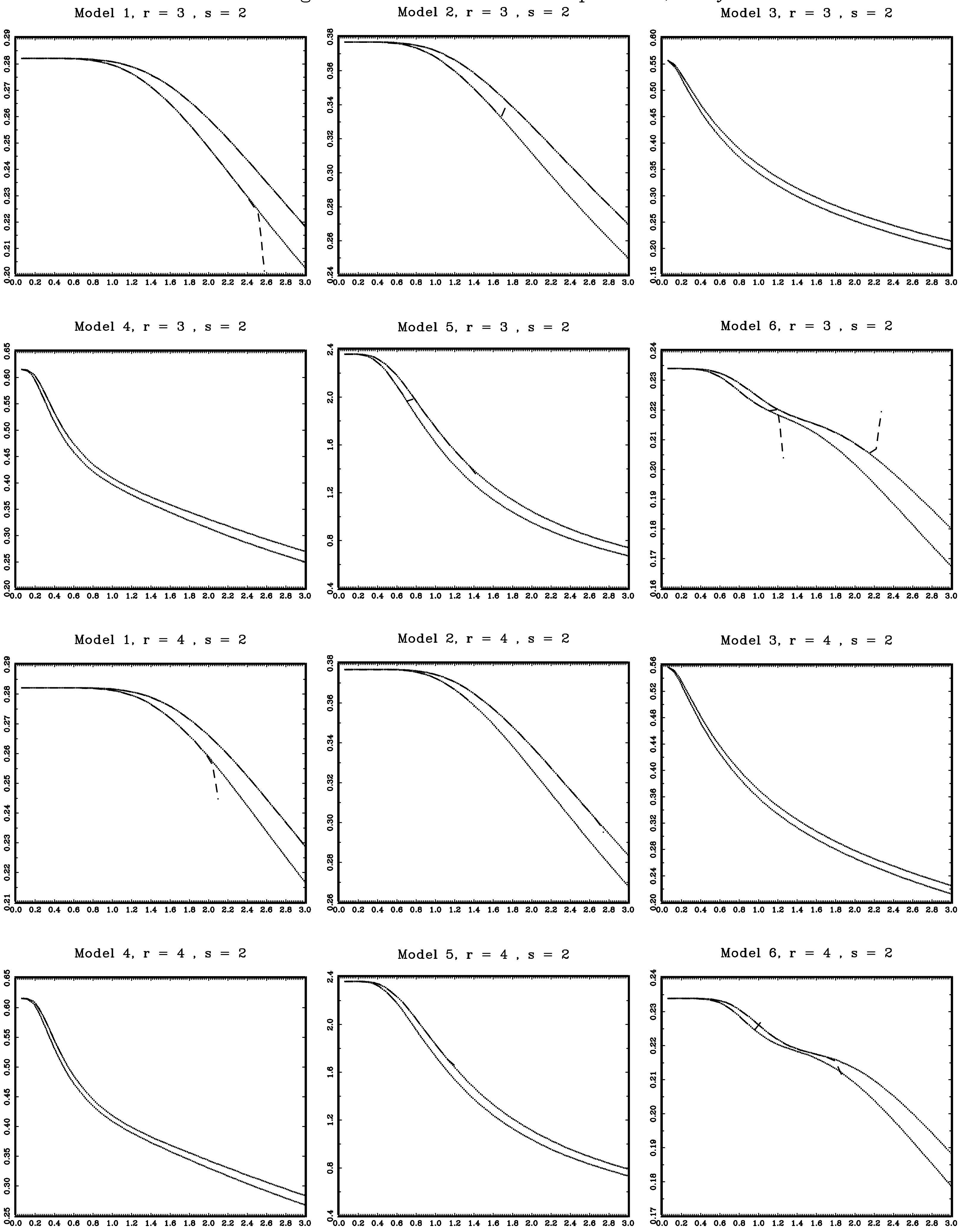
Numerical Integration Versus Series Expansion, Gaussian Kernel



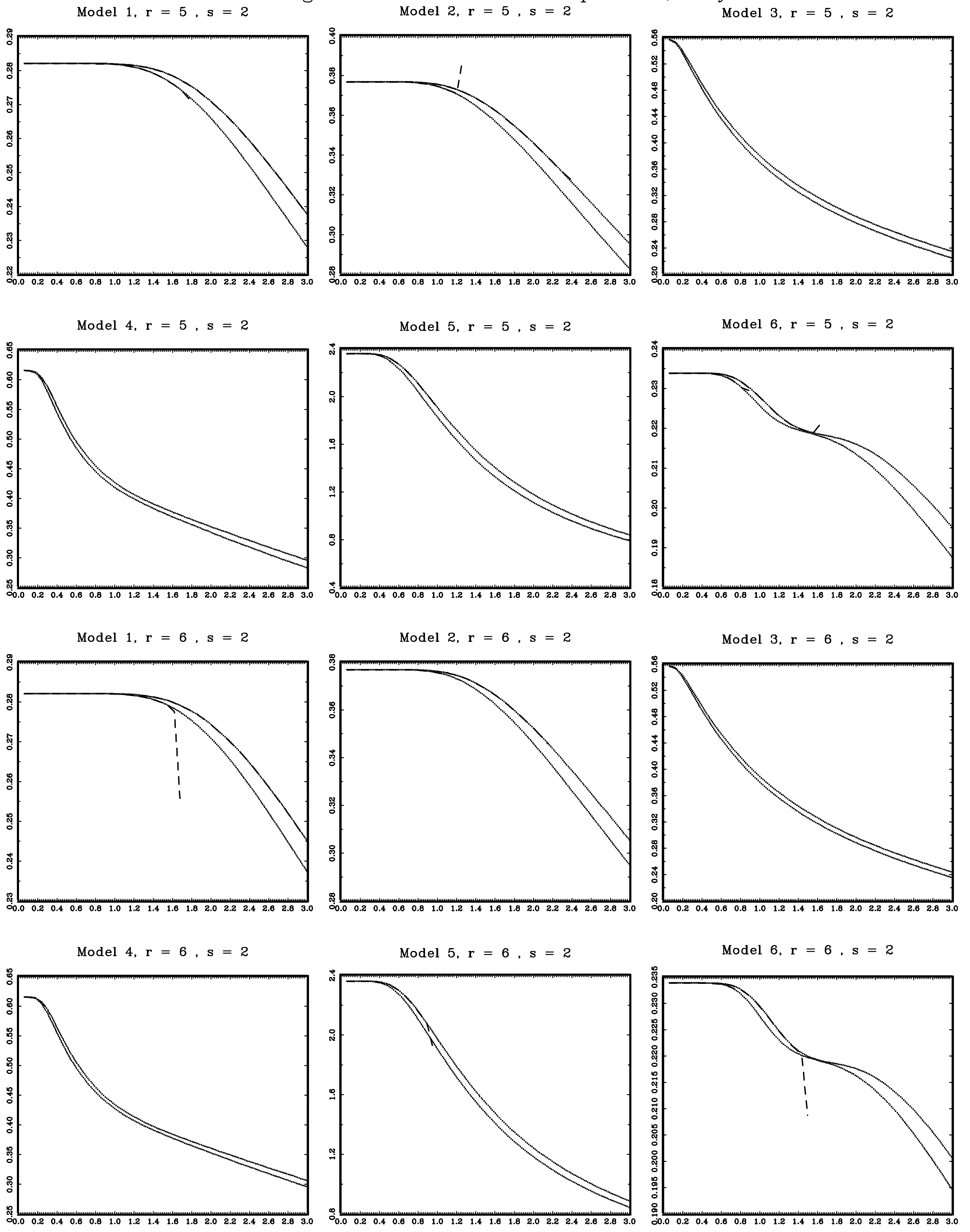
Numerical Integration Versus Series Expansion, Polynomial Kernel



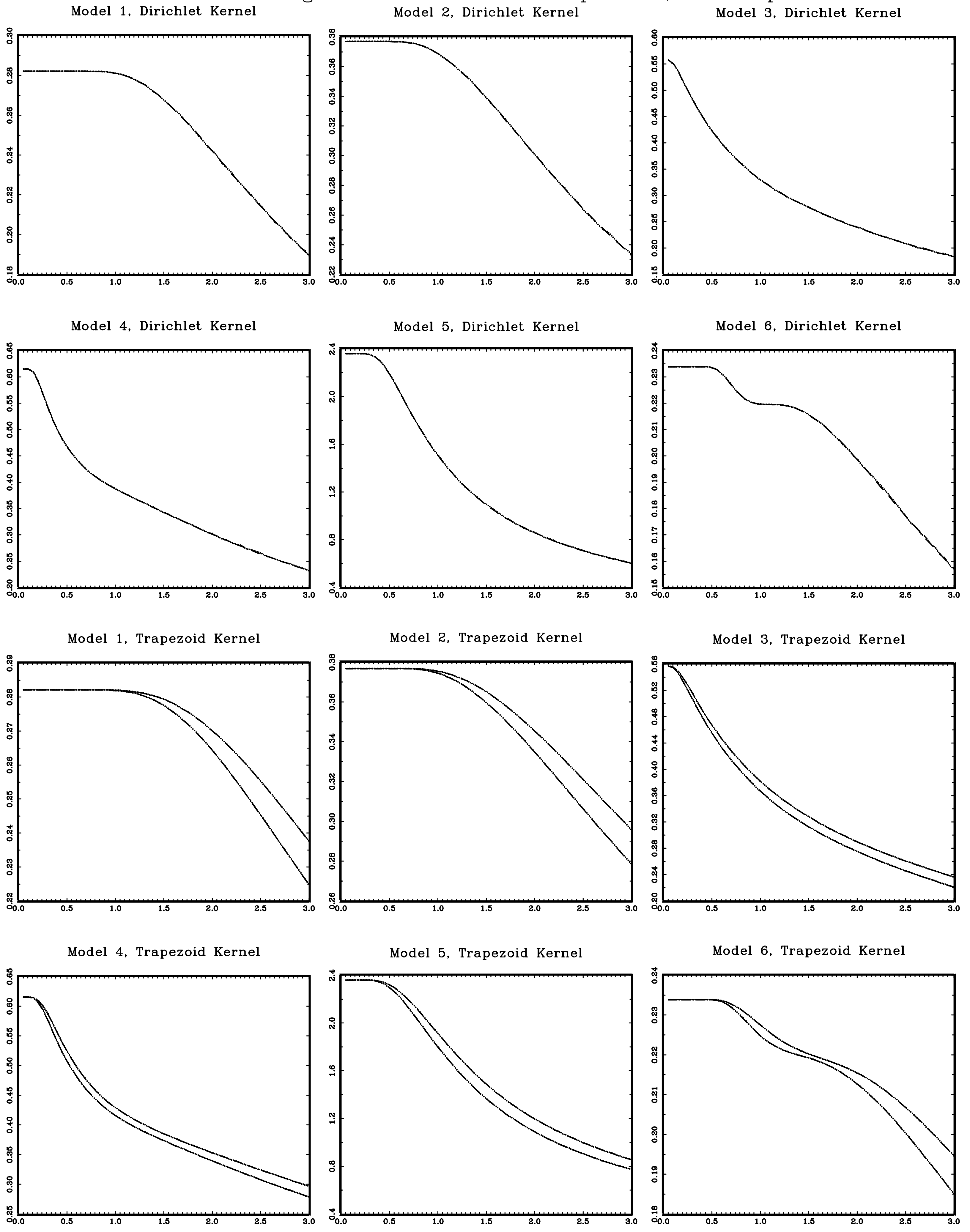
Numerical Integration Versus Series Expansion, Polynomial Kernel



Numerical Integration Versus Series Expansion, Polynomial Kernel



Numerical Integration Versus Series Expansion, Flat-Top Kernels

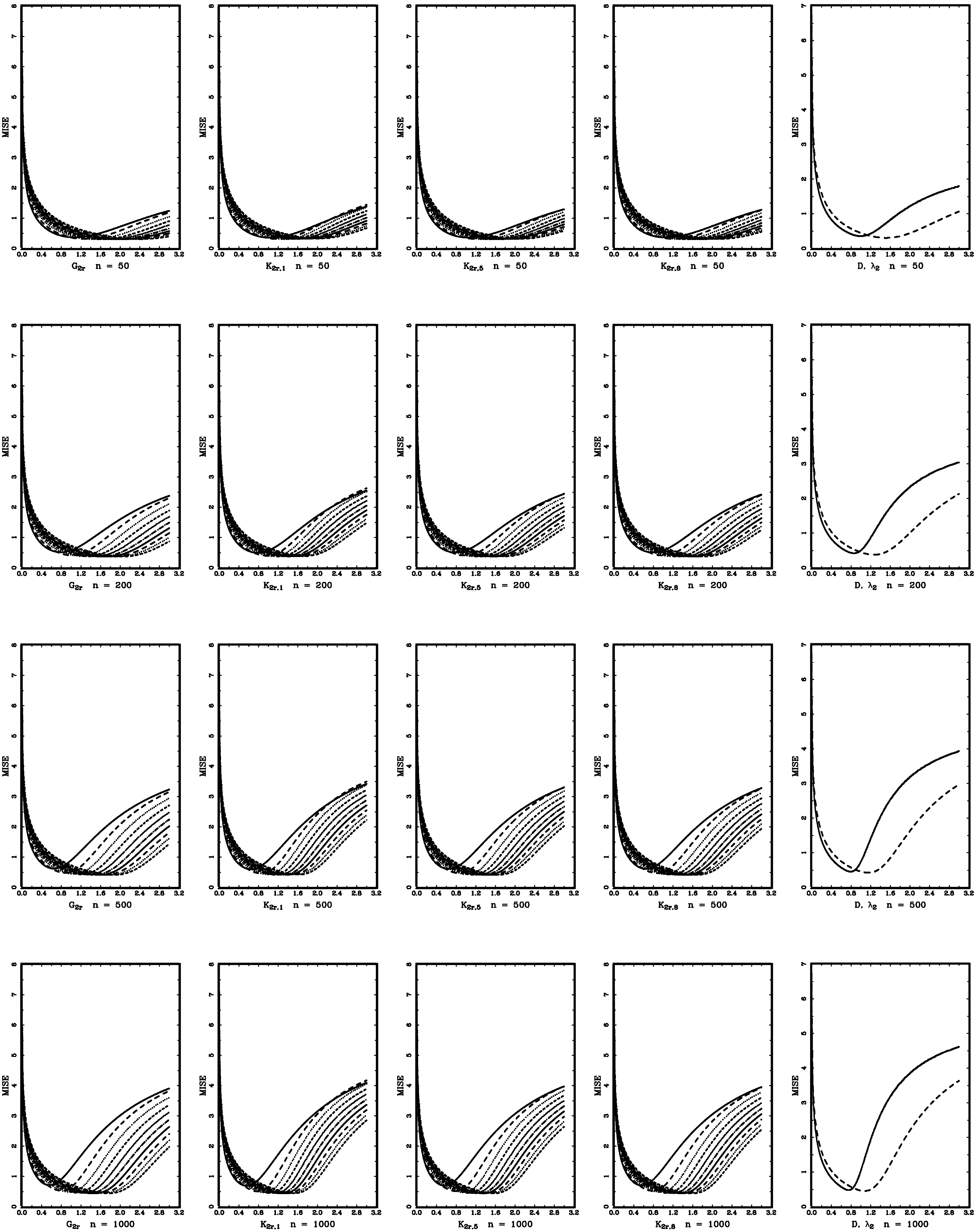


2 MISE(h)

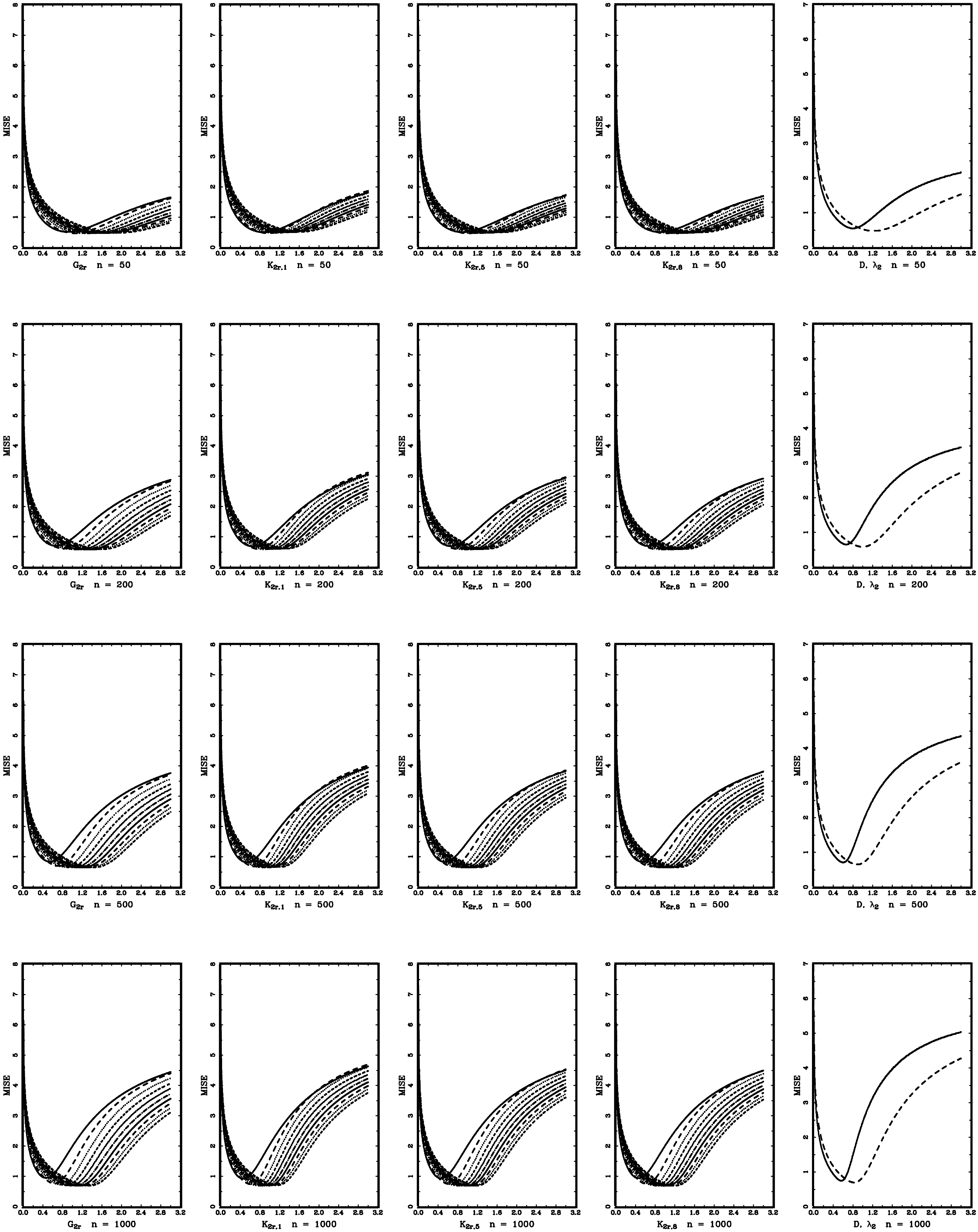
Given the integrals $I_1(h, K, f)$ and $I_2(h, K, f)$, we calculated $MISE_n(h, K, f)$ for each density f, kernel K, and a grid of values of the bandwidth h. The goal is to obtain the minimizing value $MISE_n^*(K, f) = MISE_n(h, K, f)$, so it is necessary for the grid to contain the global minimum. The following ten figures show plots of $\ln MISE_n(h, K, f)$ as a function of h for the first ten Marron-Wand densities and for the kernels used in the subsequent analysis. Each page is for one Marron-Wand density, and each of the four rows for the sample sizes $n = 50$, 200, 500, and 1000. The first column is for the Gaussian densities, and ten lines in each panel are plotted, one for each r from 1 to ten. The second, third and fourth columns are for $K_{2r,1}$, $K_{2r,5}$, and $K_{2r,8}$, again with each panel plotting $MISE$ for the then values of r. The fifth column is for the kernels D and λ_2, so only two lines are displayed.

Looking through the ten pages, we can see that most $MISE$ functions appear to have a single local minima, so we can be confident that the range of h includes the global minimum. In some cases (such as Model 3) the $MISE$ function is quite peaked as a function of h, indicating that for these models obtaining the correct bandwidth is critical. For some cases (such as Models 6 and 10) we confirm the finding of Marron and Wand (1992) that the $MISE$ function can have multiple local minima in h when the sample size n is small. This is one reason we have used grid search to minimize $MISE$ over h.

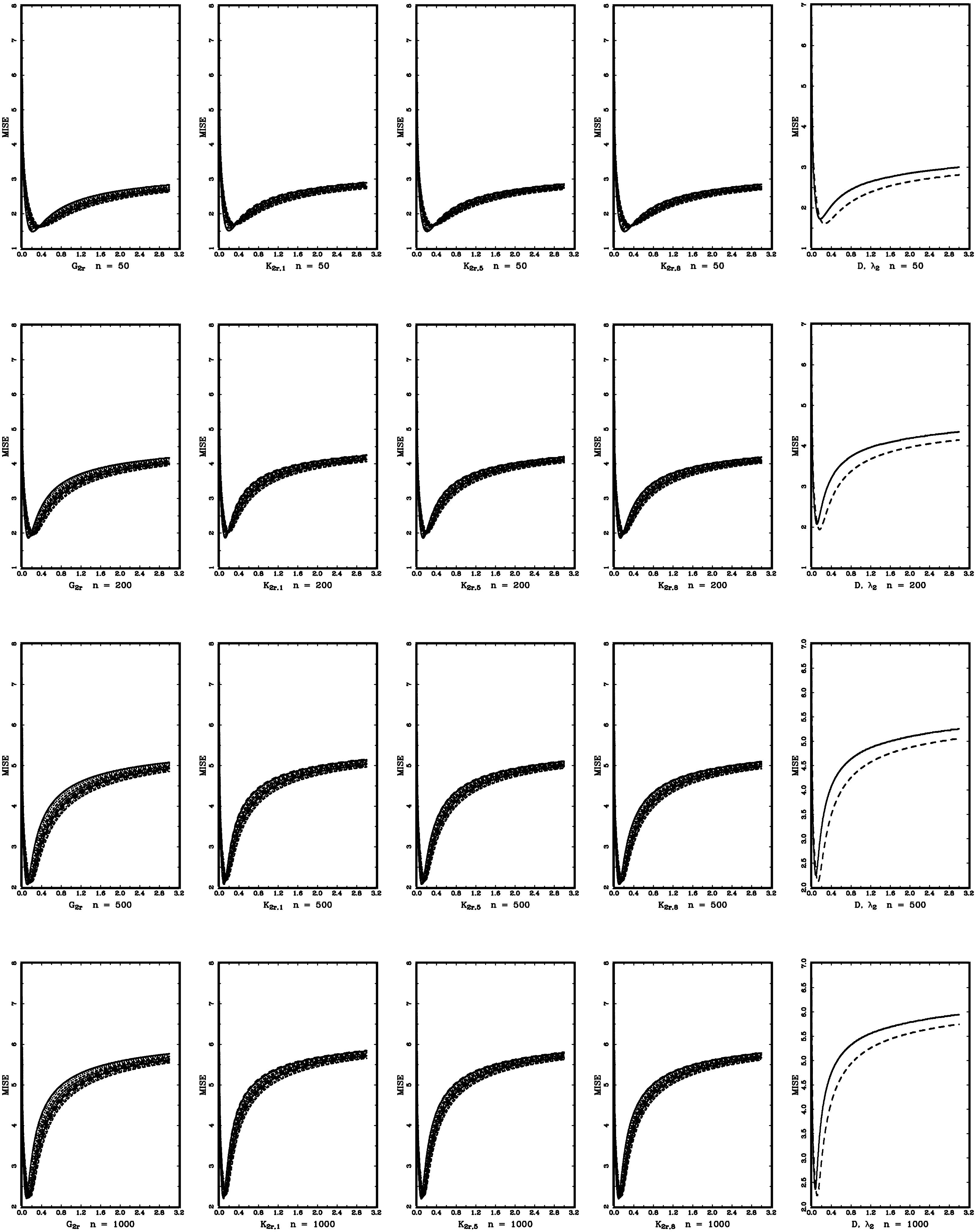
MISE(h), Density #1



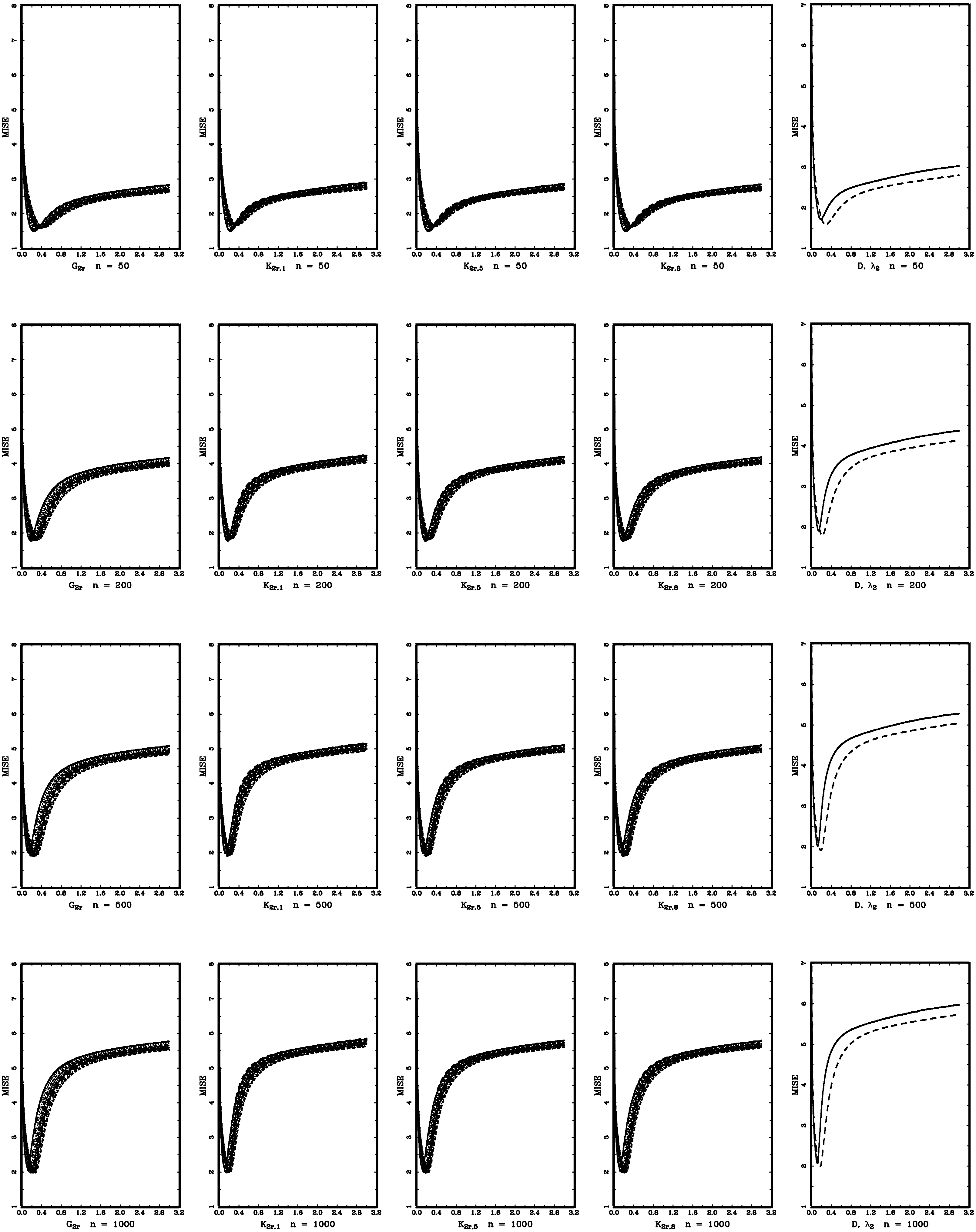
MISE(h), Density #2



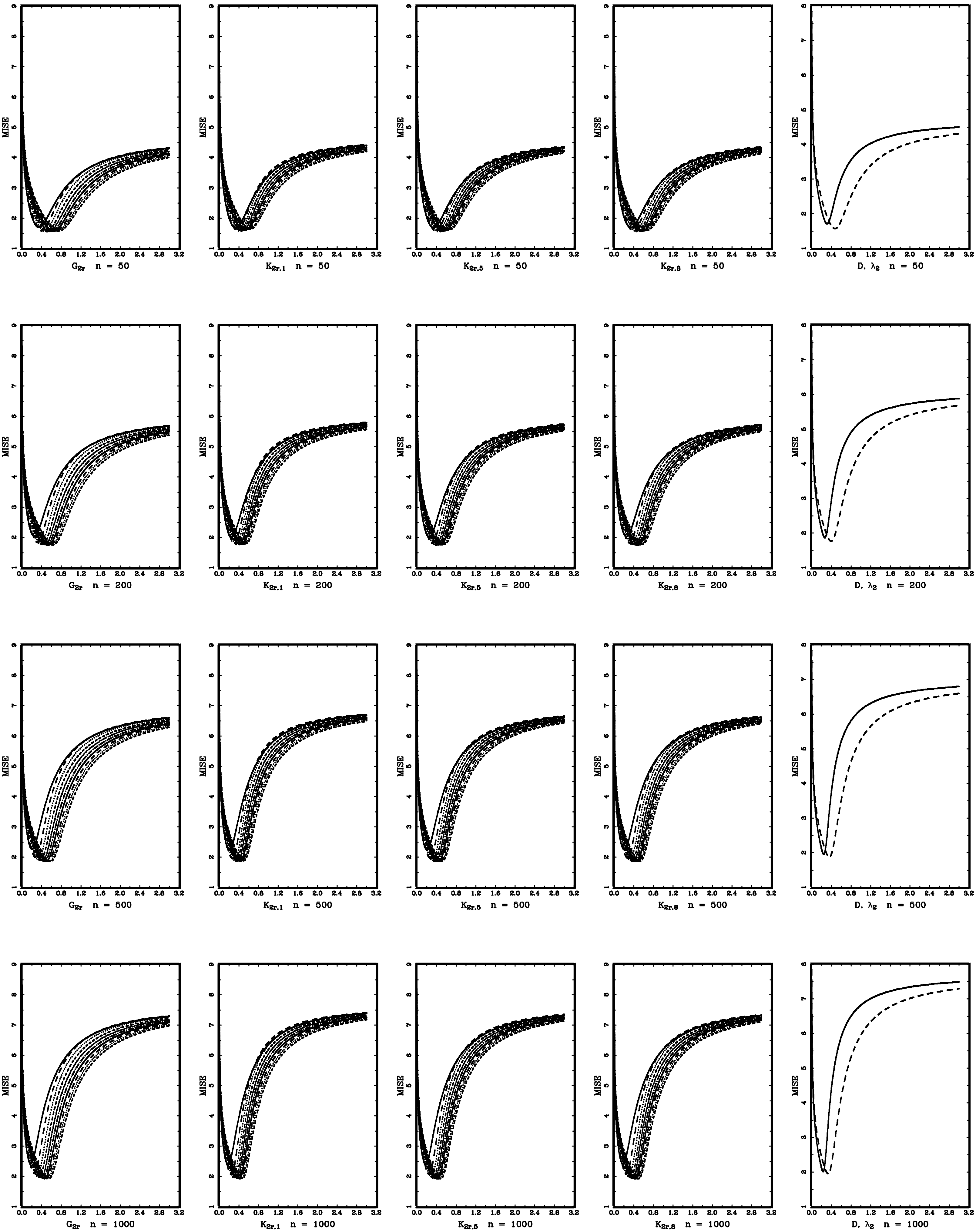
MISE(h), Density #3



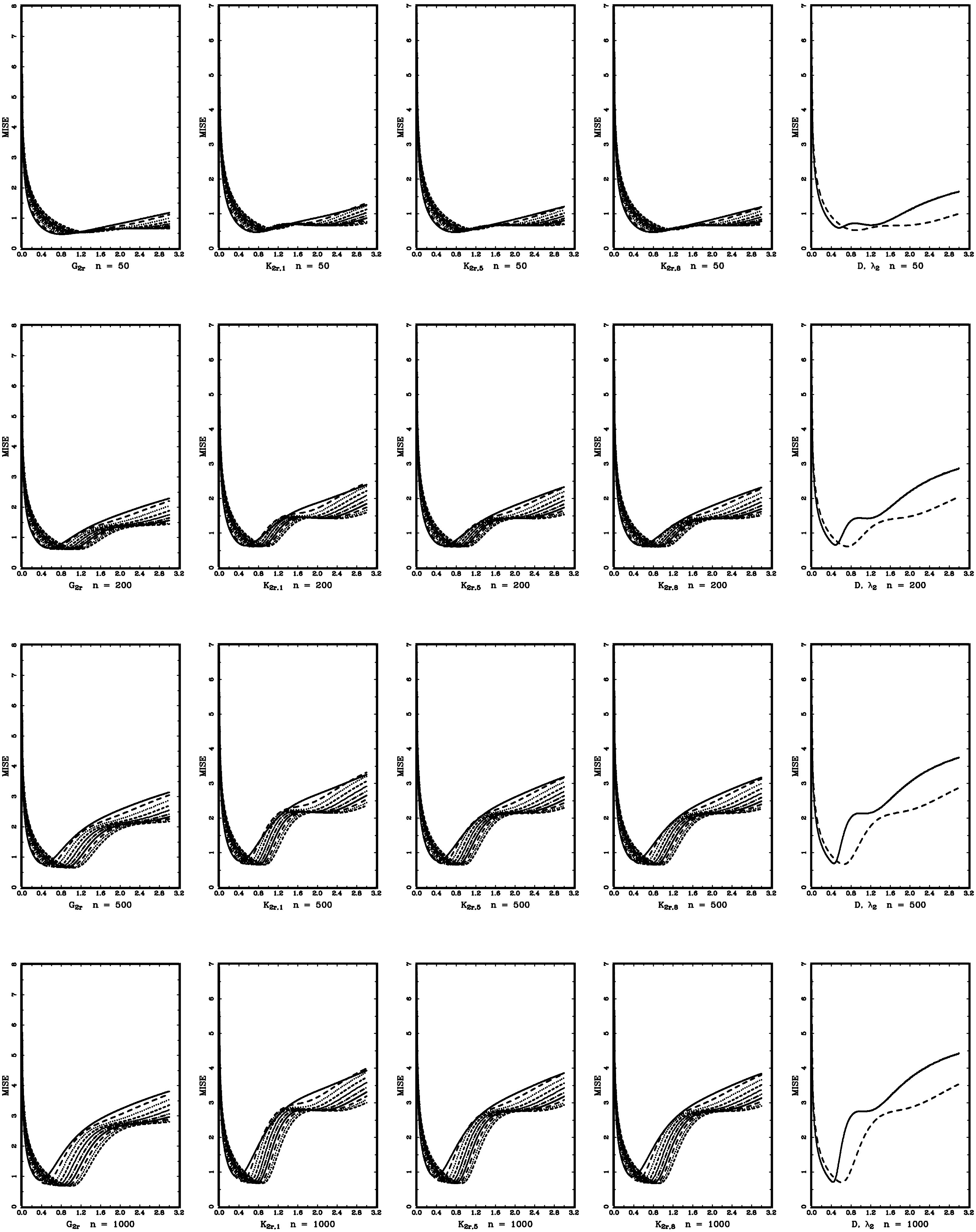
MISE(h), Density #4



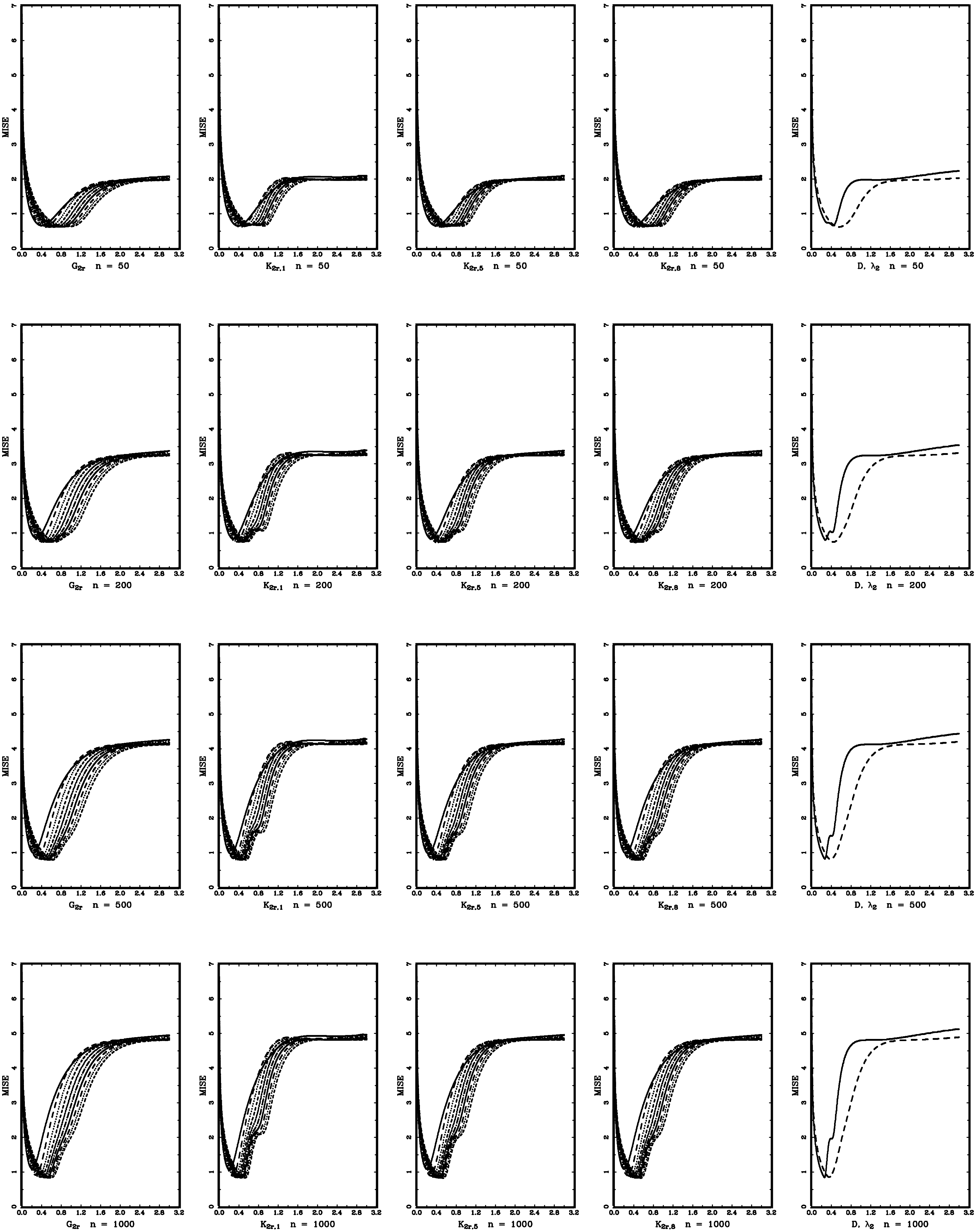
MISE(h), Density #5



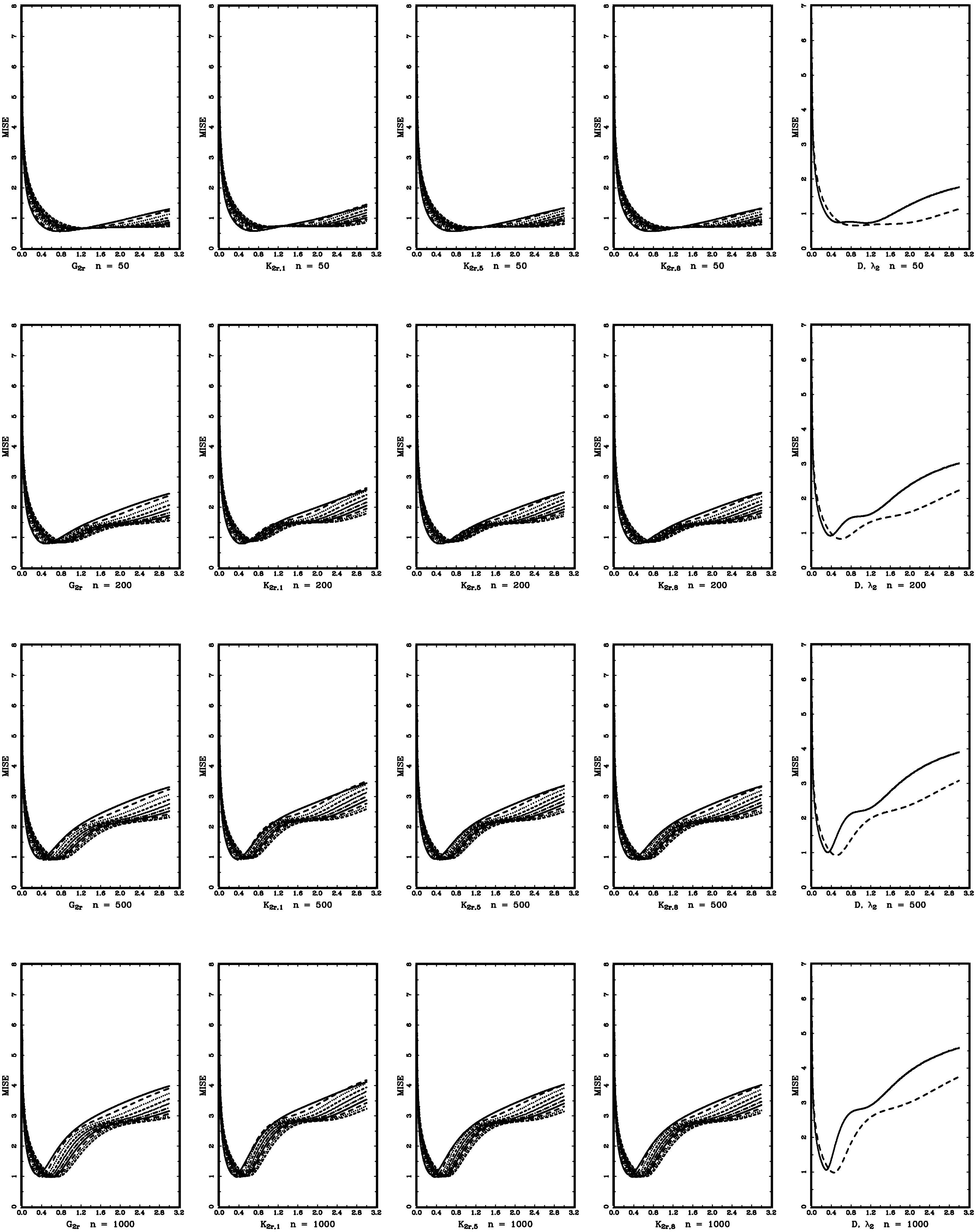
MISE(h), Density #6



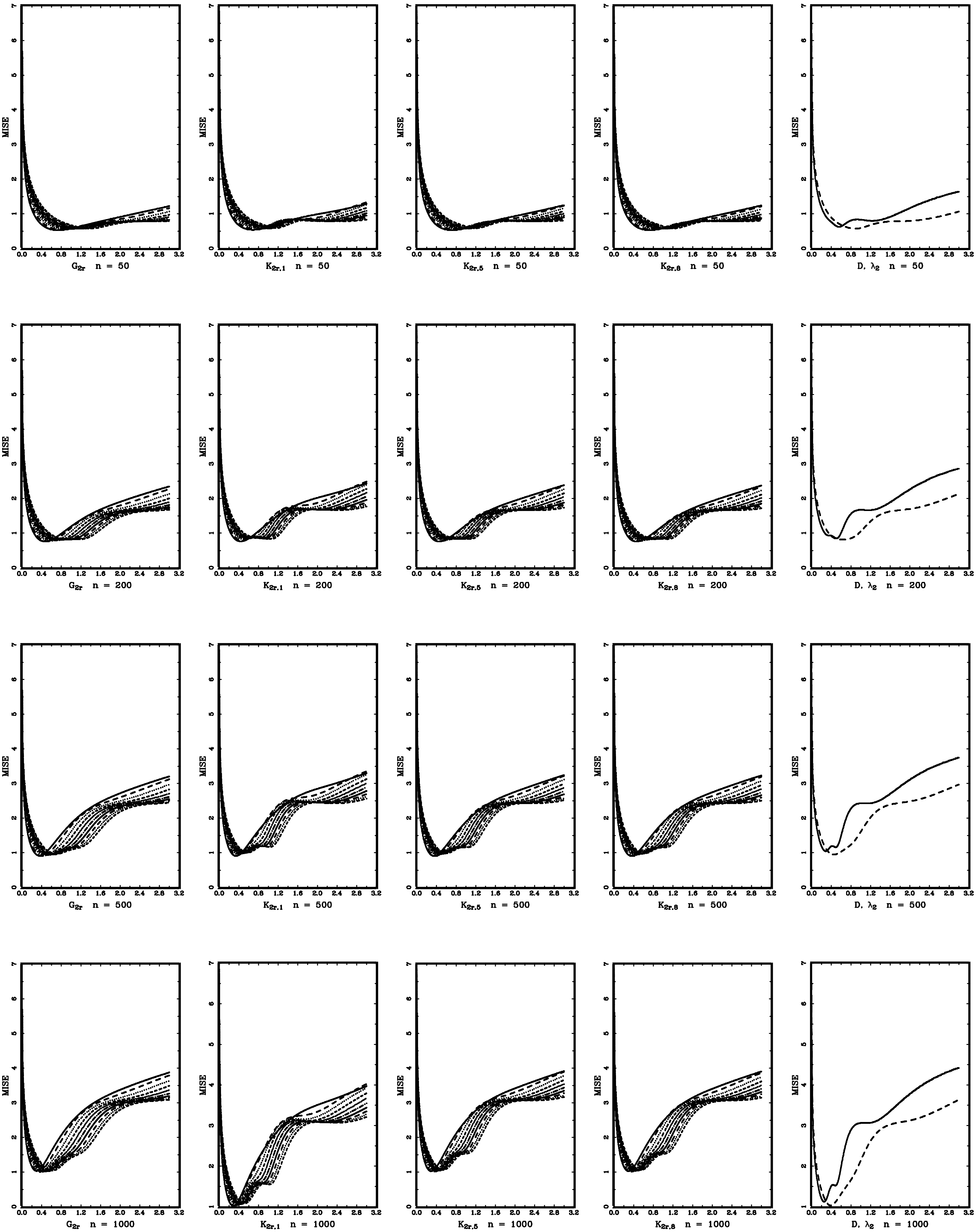
MISE(h), Density #7



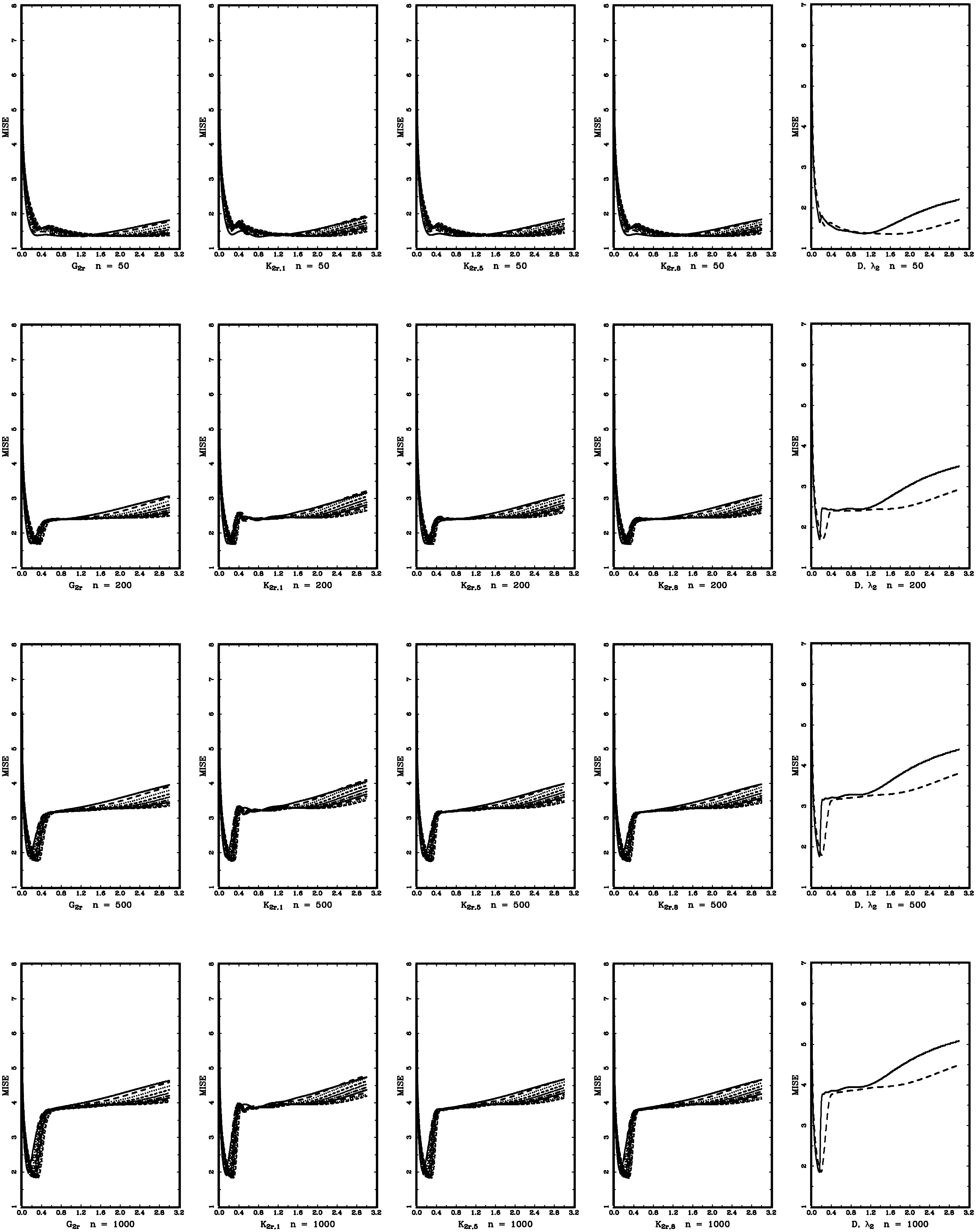
MISE(h), Density #8



MISE(h), Density #9

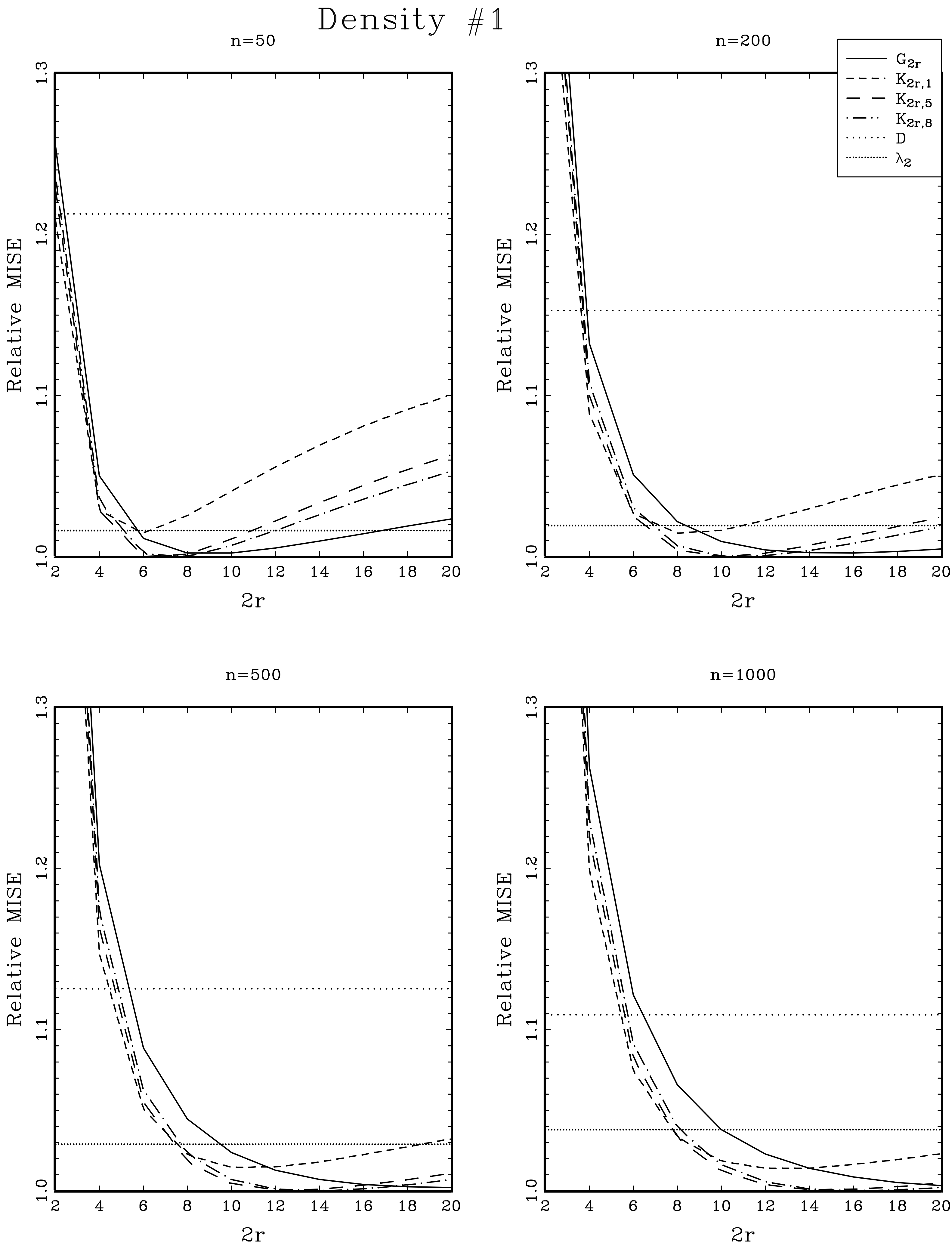


MISE(h), Density #10

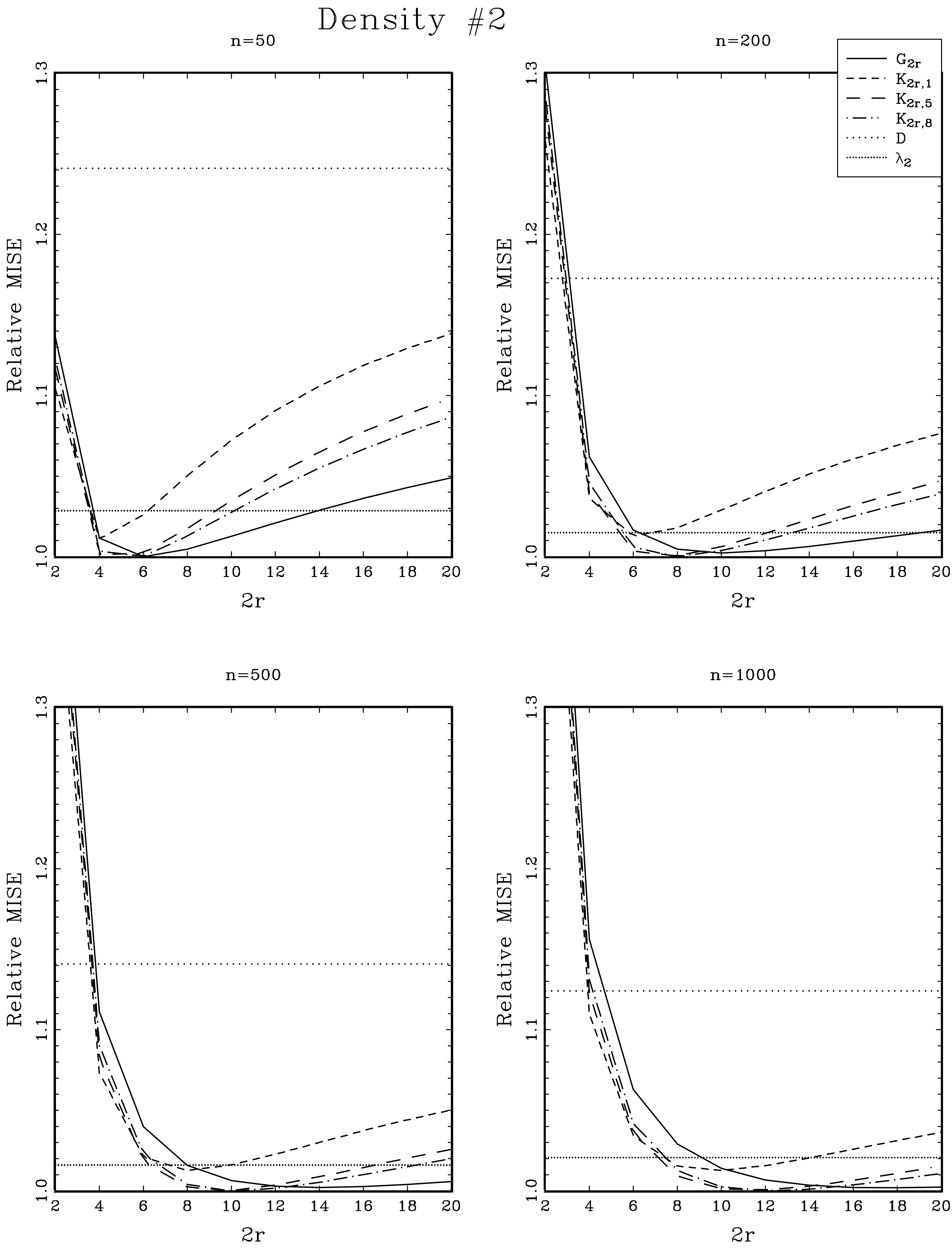


3 MISE(K,f)

For each Marron-Wand density f and kernel K, $MISE_n(K, f)$ was calculated. The following pages show plots of $MISE_n$ as functions of kernel order $2r$. In each case $MISE$ is scaled relative to the best achieving kernel for that panel, so that the y-axis measures the proportional deviation of $MISE_n$ from the optimum. Plotted are the Gaussian kernel G_{2r} and the polynomial kernels $K_{2r,1}$, $K_{2r,5}$, and $K_{2r,8}$. Also shown are the $MISE_n^*$ achieved by the kernels D and λ_2. Since these are not functions of r we have drawn them as flat lines.

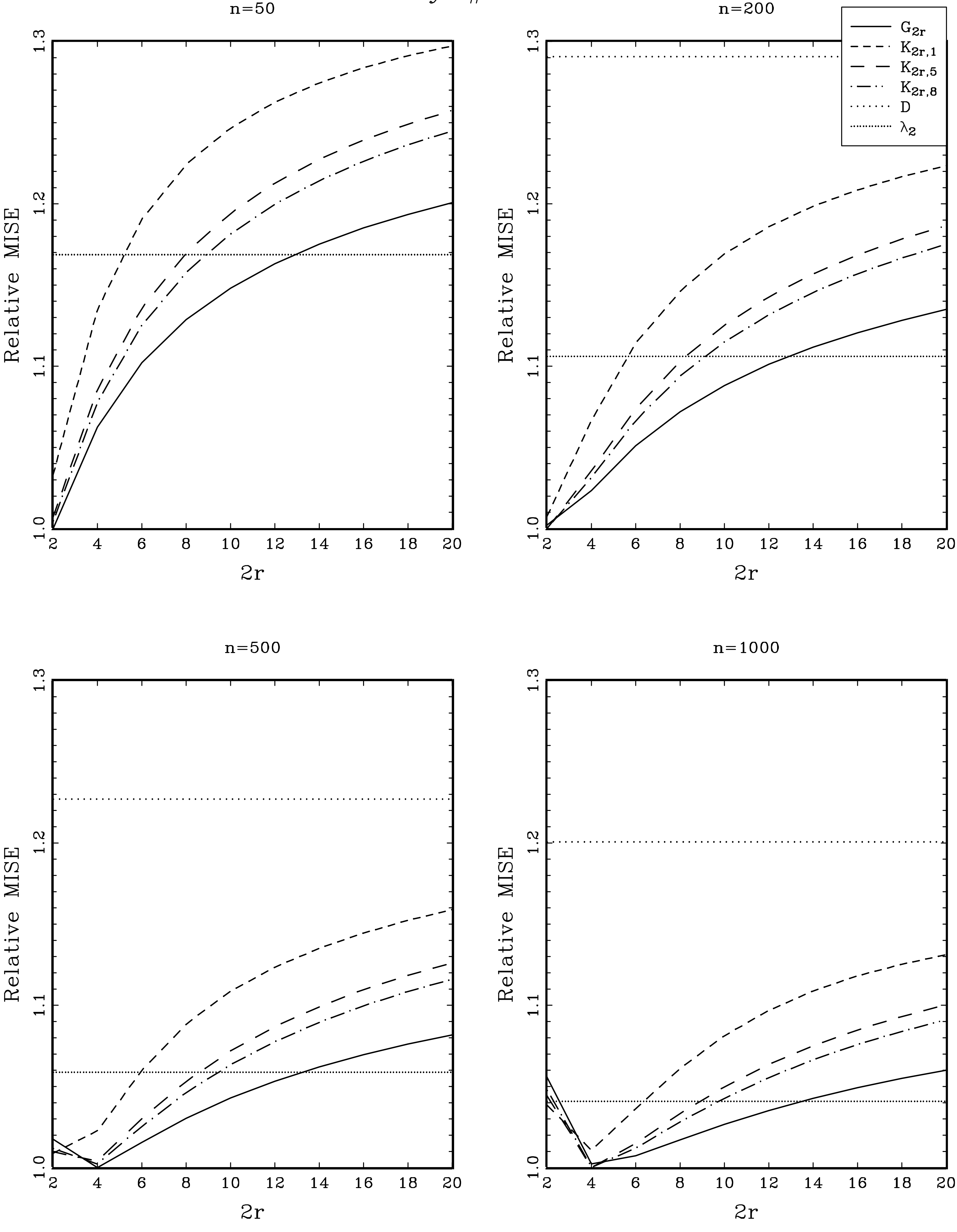


Density #1
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2



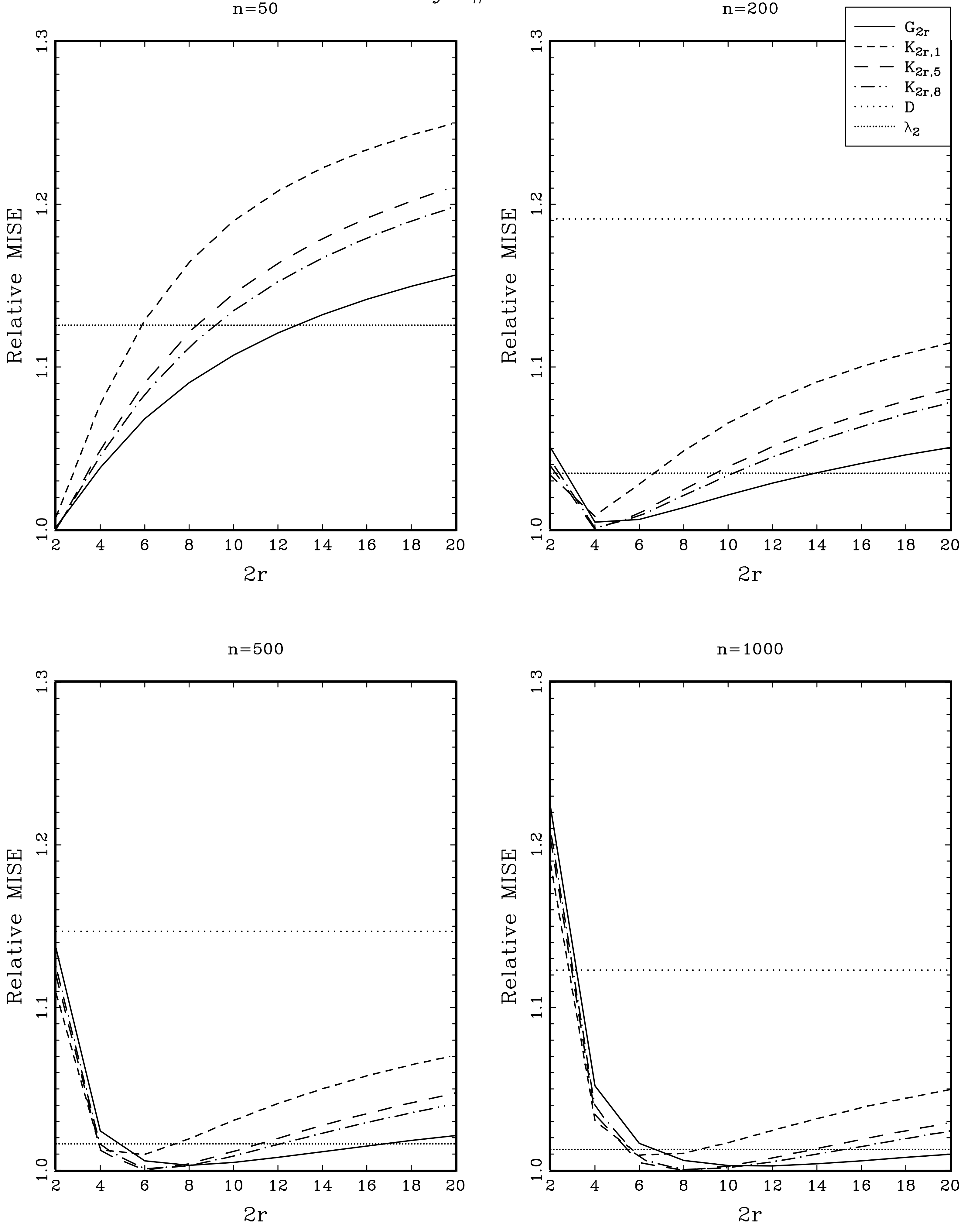
Density #2
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

Density #3

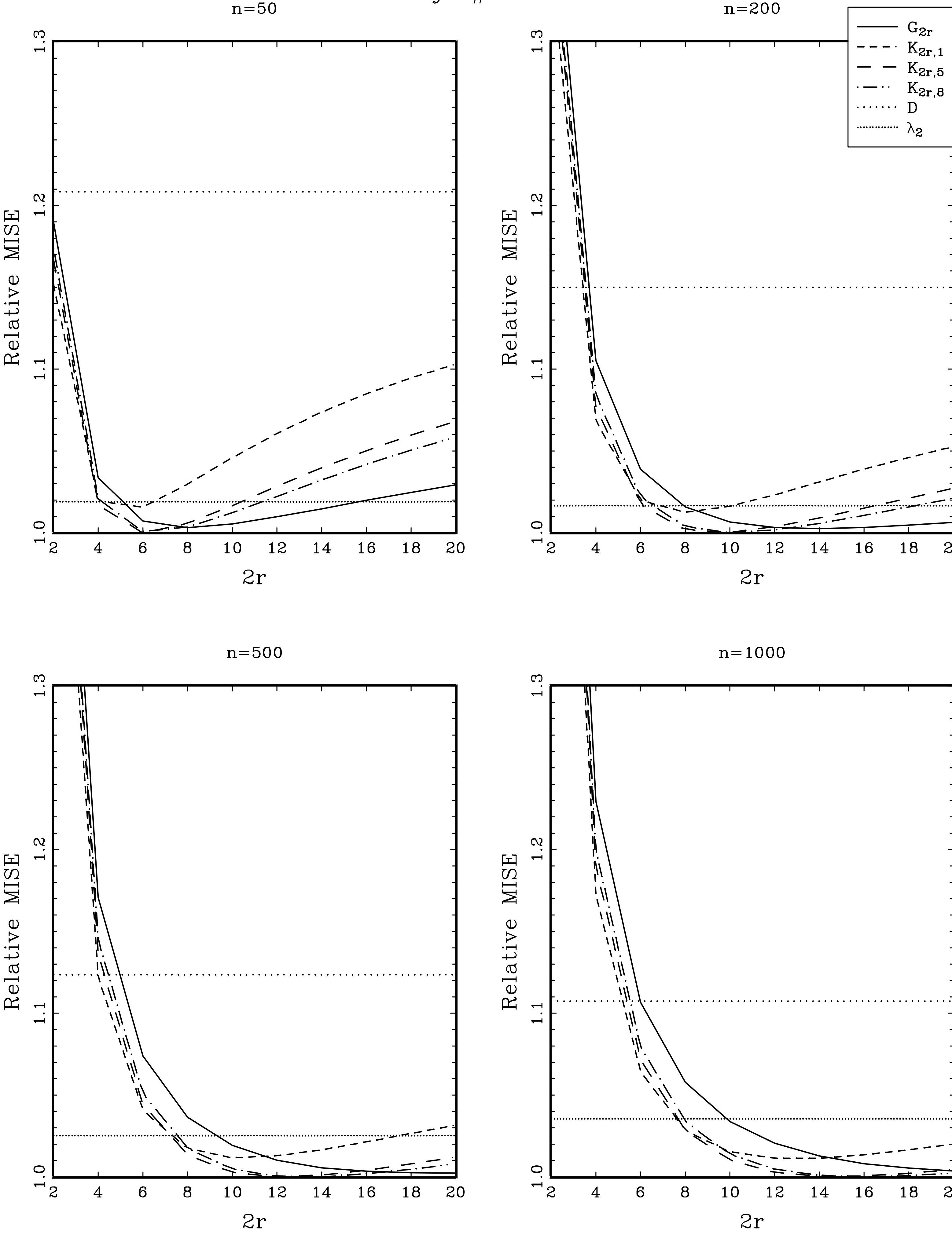


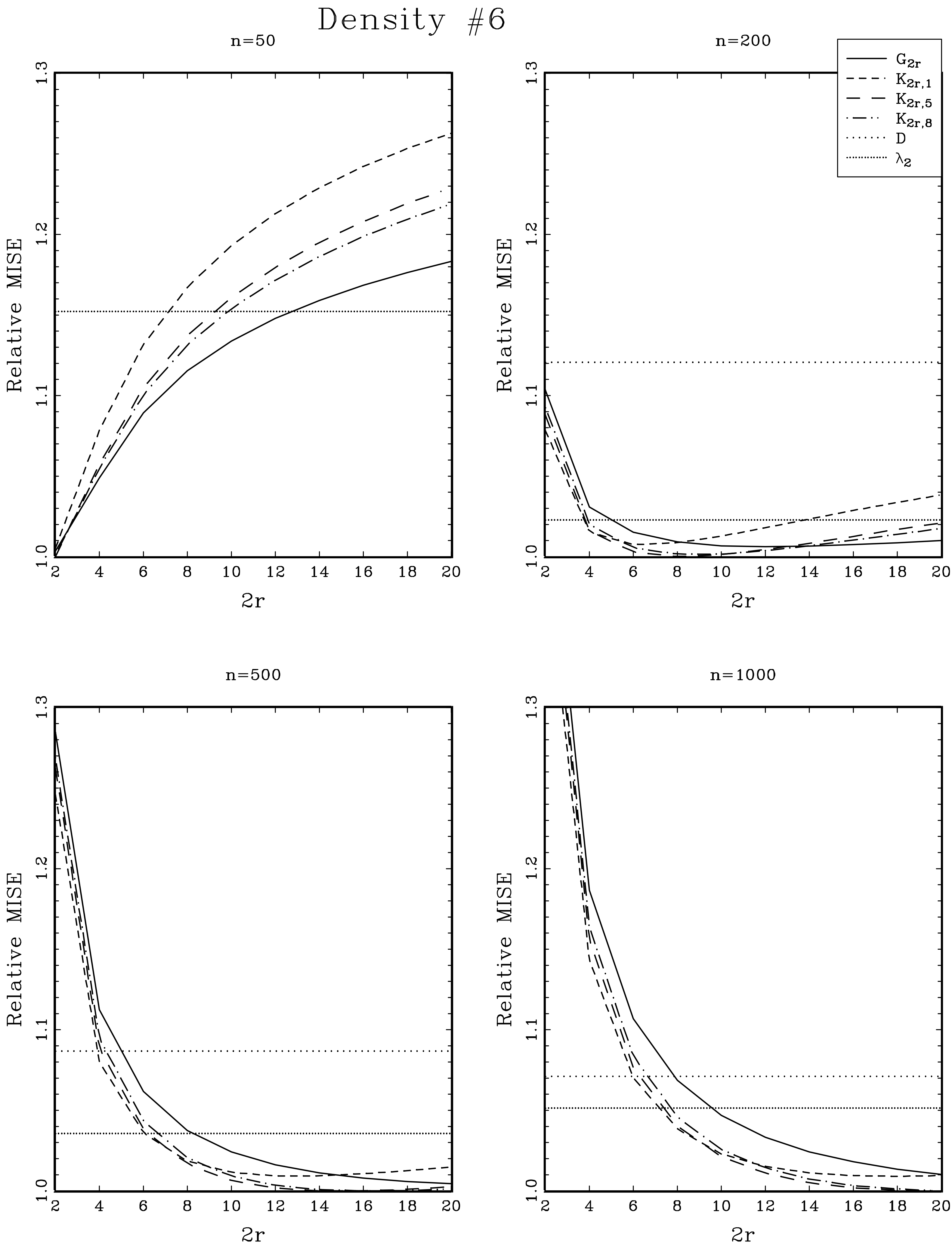
n=50
n=200
n=500
n=1000
Relative MISE
2r
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

Density #4

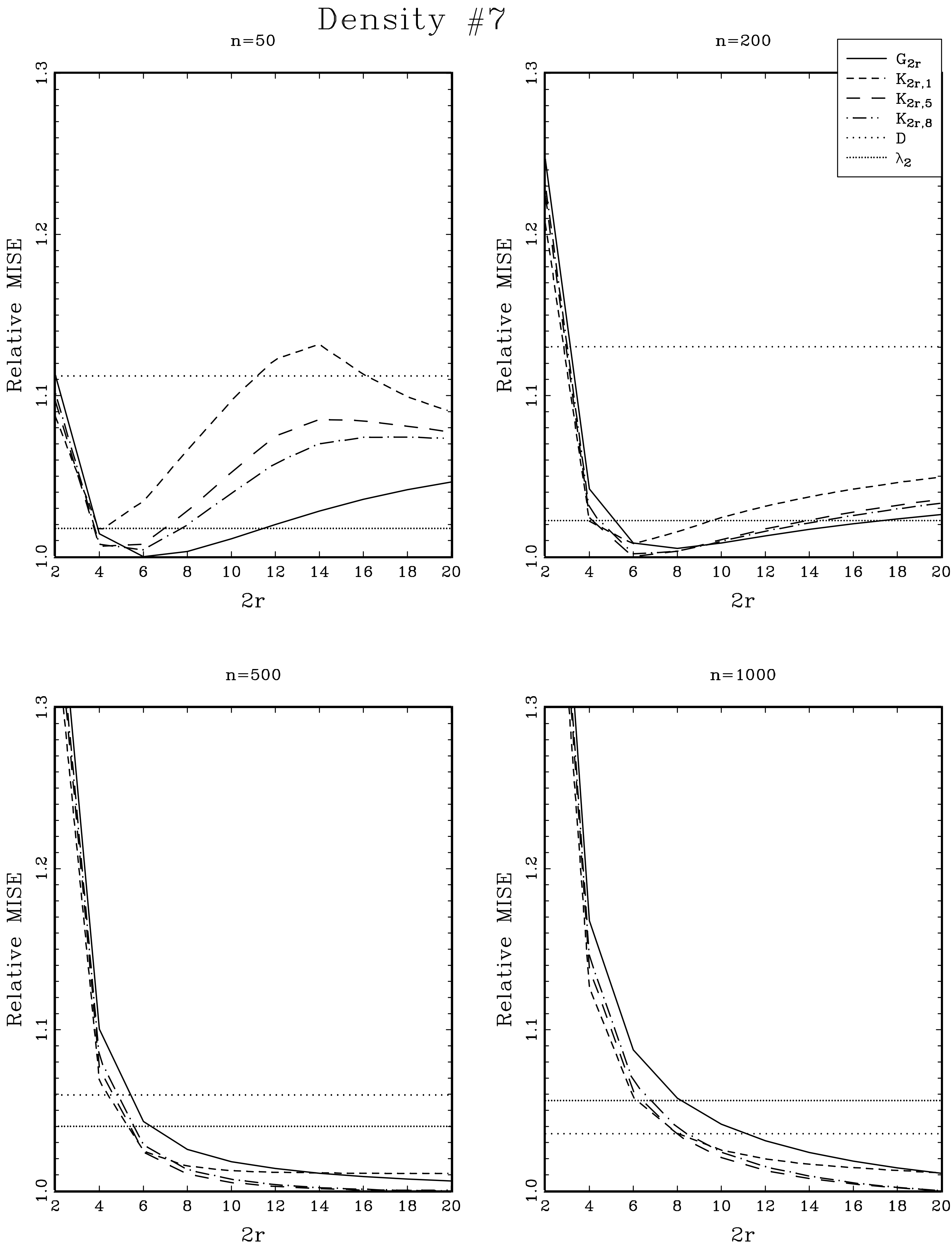

n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

Density #5

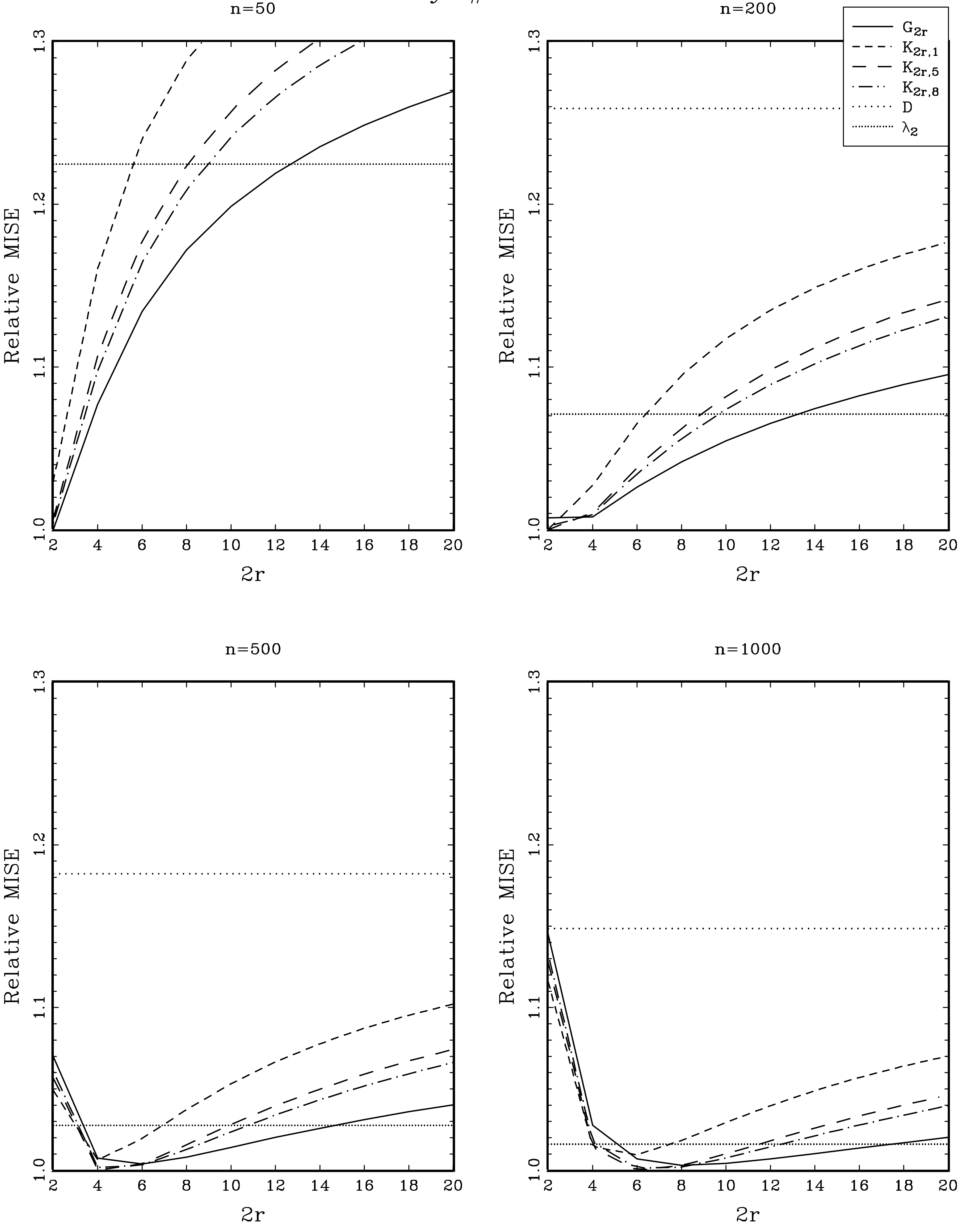

n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2



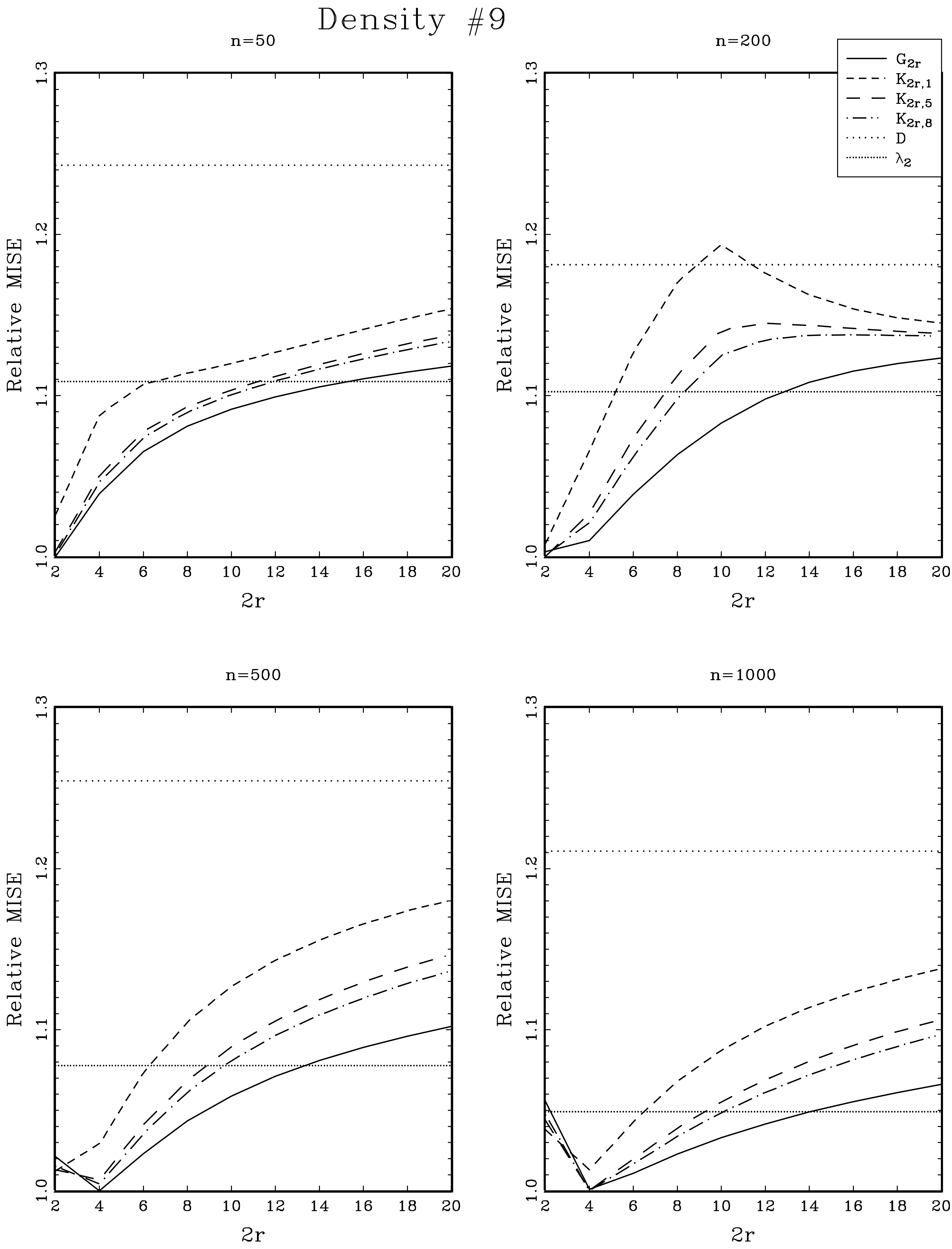
Density #6
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G2r
K2r,1
K2r,5
K2r,8
D
λ2

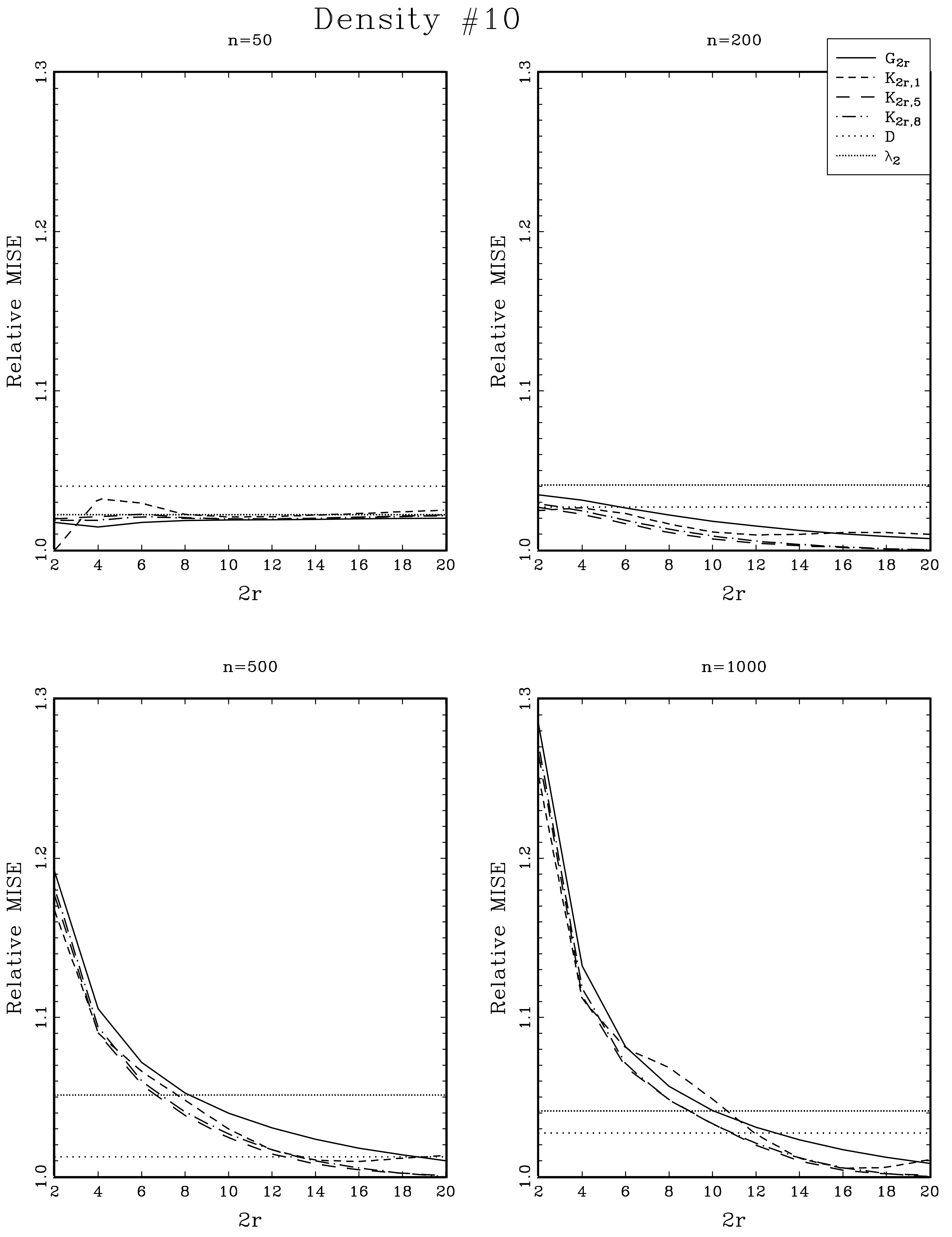

Density #7
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

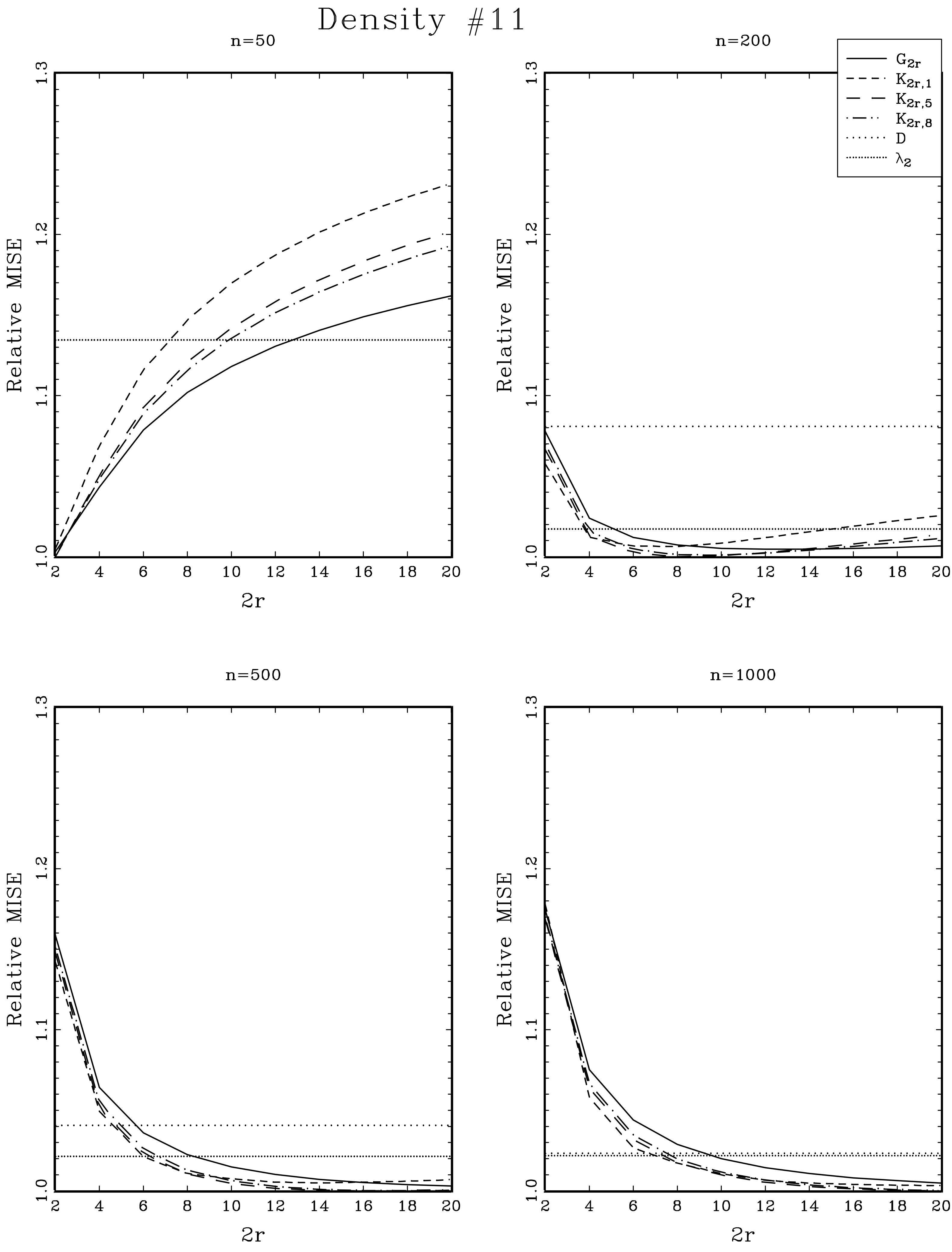
Density #8



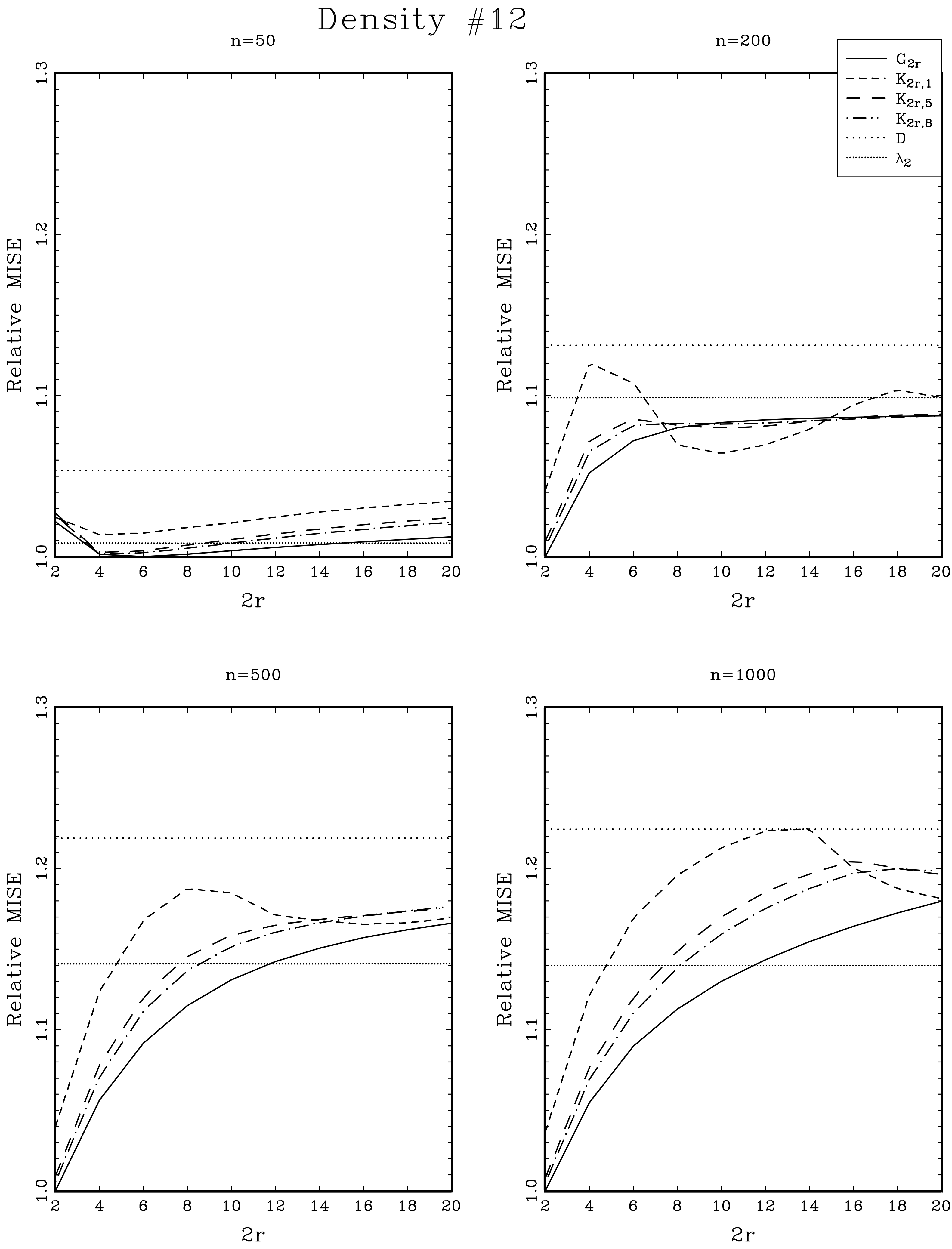
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

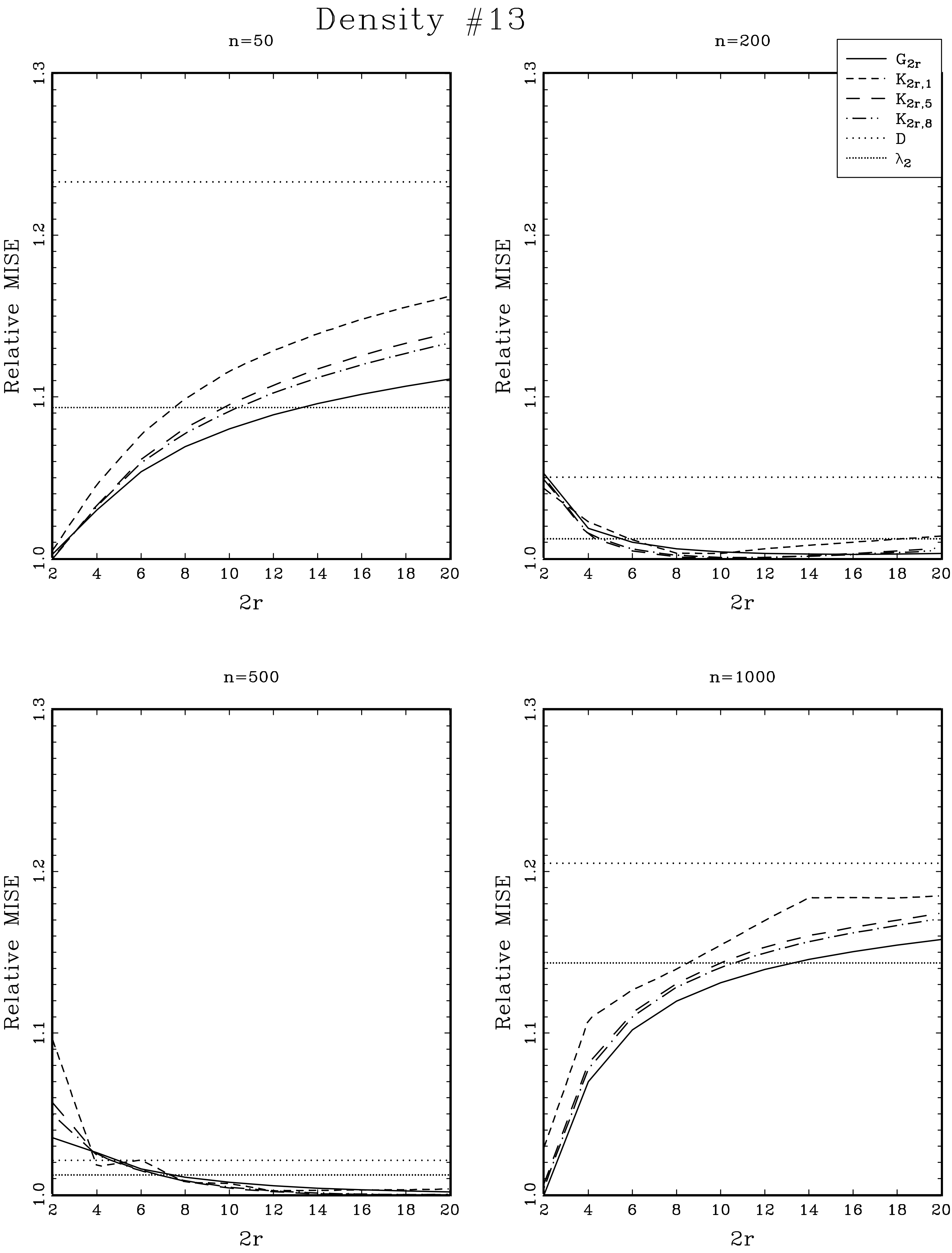

Density #9
n=50
n=200
n=500
n=1000
Relative MISE
2r
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

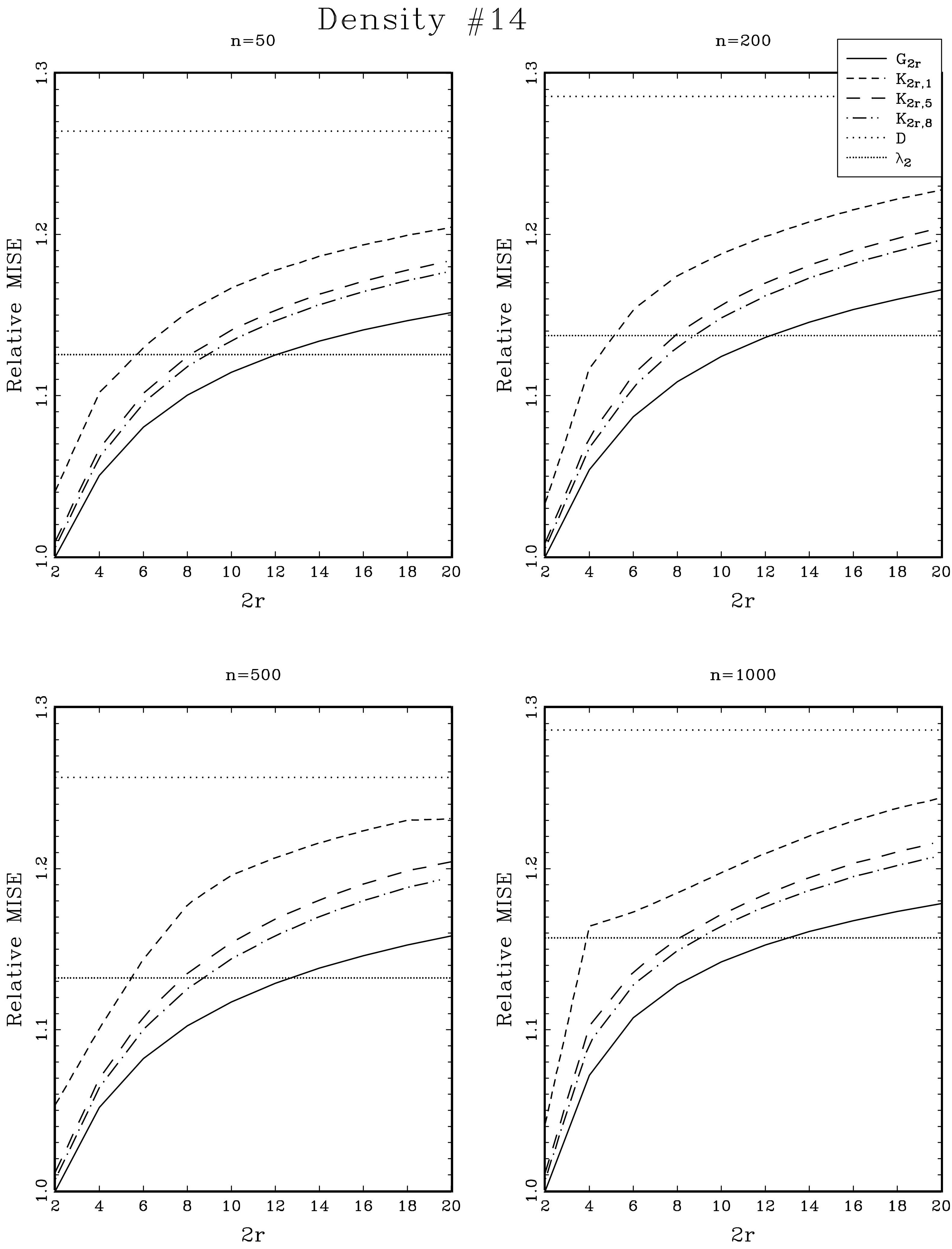

Density #10
n=50
n=200
n=500
n=1000
Relative MISE
2r
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

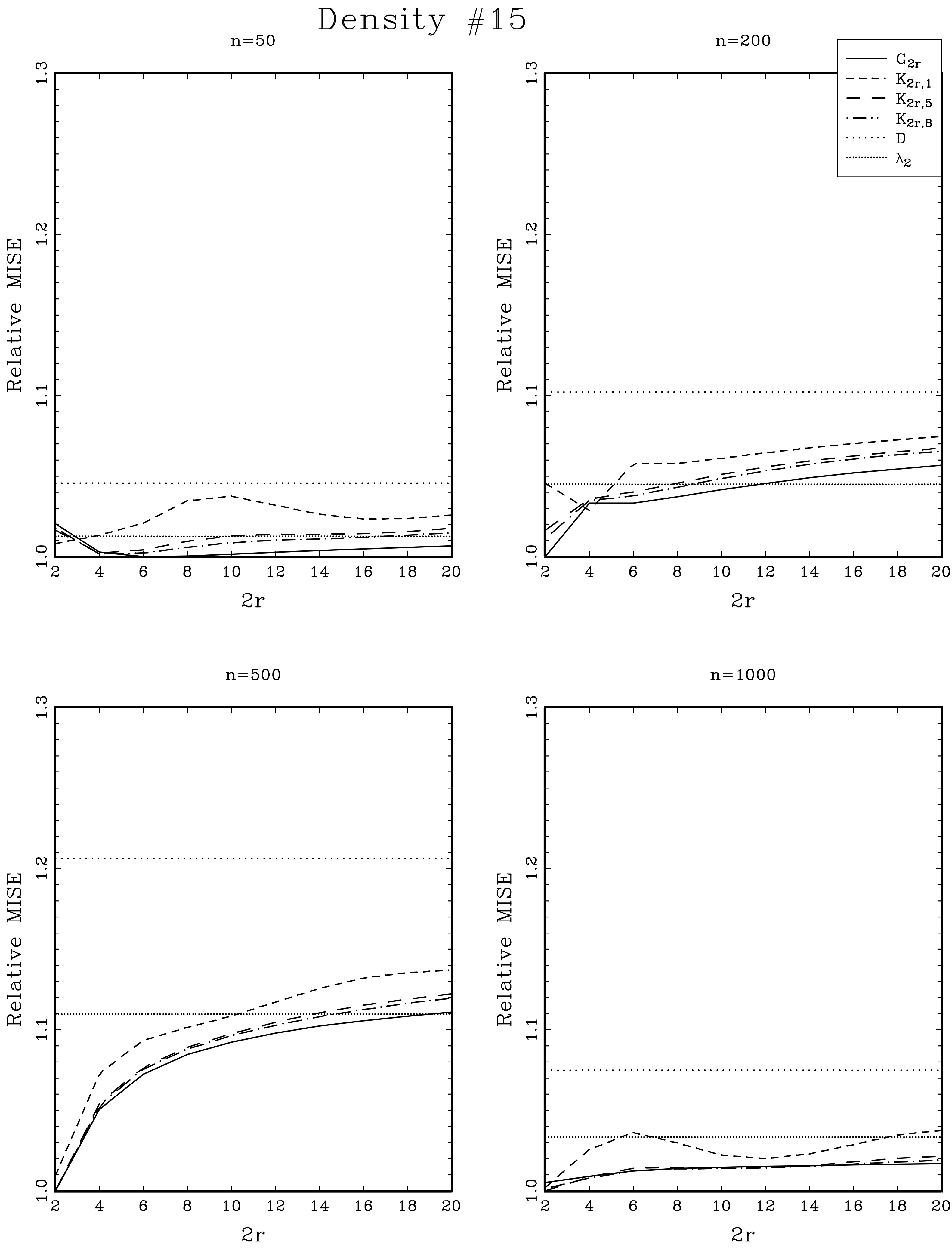


Density #11
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2


Density #12
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2


Density #13
n=50
n=200
n=500
n=1000
Relative MISE
2r
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20


Density #14
n=50
n=200
n=500
n=1000
Relative MISE
2r
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2



Density #15
n=50
n=200
n=500
n=1000
Relative MISE
2r
1.0
1.1
1.2
1.3
2
4
6
8
10
12
14
16
18
20
G_{2r}
$K_{2r,1}$
$K_{2r,5}$
$K_{2r,8}$
D
λ_2

4 Regret

Figure 4 of the paper displays the regret

$$\text{Regret}_n\,(f,K) = \ln MISE_n(K,f) - \min_{K\in\mathcal{K}} \ln MISE_n(K,f).$$

for the Gaussian kernel G_{2r}. We now present an analog for the Polynomial kernel $K_{2r,8}$. Each of the ten light lines are the regret for one of the Marron-Wand densities f, and display the regret as a function of kernel order $2r$. The heavy solid line is the maximum across the models. The lowest value of the solid line is the minimax regret.

Regret Using Polynomial Kernel