Business, Economic and Financial Data — Exam D (Advanced & Long Version)

Comprehensive exam — 2 hours total time — deep analysis required

1) Comprehensive model integration task

A firm provides quarterly data on product sales (Y), marketing spend (X1), competitor activity index (X2), and consumer sentiment (X3). You estimate multiple models to forecast next year's sales.

a) Specify a multiple regression model and interpret the coefficients conceptually.

b) Explain how to test for autocorrelation, multicollinearity, and heteroskedasticity (mention tests and interpretation).

c) Propose a dynamic model (ARIMAX or state-space form) to capture serial dependence and exogenous effects.

d) Discuss model validation strategy (train/test split, residual diagnostics, out-of-sample forecast evaluation).

2) Time series structural analysis and forecasting

Given the monthly log returns of a financial index, describe the difference between modeling with ARIMA, ARCH/GARCH, and state-space approaches.

a) Write down a GARCH(1,1) specification and interpret parameters α, β, and ω.

b) Compare its assumptions with those of exponential smoothing and ARIMA.

c) Explain how you would test for volatility clustering and model adequacy in R (include relevant R functions).

3) Nonlinear diffusion and forecasting of innovation adoption

You observe adoption data of a fintech product. Using R, you estimate a Bass and a Generalized Bass Model (GBM).

a) Write both model equations and define parameters (p, q, m, δ).

b) Given the parameter estimates below, interpret their economic meaning and discuss potential overfitting or misspecification.

c) Suggest how you would include external shocks or marketing effects dynamically.

Model	p	q	m	δ	R^2
Bass	0.018	0.135	100000	–	0.92
Gen. Bass	0.015	0.129	100000	0.38	0.97

4) Tree-based ensemble models for financial data

You are asked to predict quarterly profit growth using gradient boosting and random forest models.

a) Explain key differences between Bagging, Random Forests, and Gradient Boosting in terms of variance/bias reduction.

b) How would you tune hyperparameters (learning rate, number of trees, max depth) using cross-validation in R?

c) Describe variable importance measures and how to interpret them in economic context.

d) Discuss limitations of tree-based methods when dealing with small time series datasets.

5) Model comparison and communication of results

a) Suppose you obtain the following results from three competing models (ARIMA, GAM, Gradient Boosting). Interpret them critically.

b) Which model would you select and why? Consider both statistical performance and managerial interpretability.

c) Suggest how you would communicate these results to non-technical stakeholders.

Model	RMSE	MAPE	R^2	Interpretability
ARIMA(1,1,1)	245	6.4%	0.89	High
GAM	198	4.7%	0.93	Moderate
GBM	182	4.1%	0.95	Low