Context-Adaptive Inference: Bridging Statistical and Foundation Models

This manuscript (permalink) was automatically generated from AdaptInfer/context-review@95d13f2 on July 31, 2025.

Authors

- **Ben Lengerich**
  0000-0001-8690-9554 · blengerich · ben_lengerich
  Department of Statistics, University of Wisconsin-Madison · Funded by None

- **Caleb N. Ellington**
  0000-0001-7029-8023 · cnellington · probablybots
  Computational Biology Department, Carnegie Mellon University · Funded by None

- **Yue Yao**
  0009-0000-8195-3943 · YueYao-stat
  Department of Statistics, University of Wisconsin-Madison · Funded by None

✉ — Correspondence possible via GitHub Issues

Abstract

Context-adaptive inference enables models to adjust their behavior across individuals, environments, or tasks. This adaptivity may be *explicit*, through parameterized functions of context, or *implicit*, as in foundation models that respond to prompts and support in-context learning. In this review, we connect recent developments in varying-coefficient models, contextualized learning, and in-context learning. We highlight how foundation models can serve as flexible encoders of context, and how statistical methods offer structure and interpretability. We propose a unified view of context-adaptive inference and outline open challenges in developing scalable, principled, and personalized models that adapt to the complexities of real-world data.

Introduction

A convenient simplifying assumption in statistical modeling is that observations are independent and identically distributed (i.i.d.). This assumption allows us to use a single model to make predictions across all data points. But in practice, this assumption rarely holds. Data are collected across different individuals, environments, and tasks – each with their own characteristics, constraints, and dynamics.

To model this heterogeneity, a growing class of methods aim to make inference *adaptive to context*. These include varying-coefficient models in statistics, transfer and meta-learning in machine learning, and in-context learning in large foundation models. Though these approaches arise from different traditions, they share a common goal: to use contextual information – whether covariates, environments, or support sets – to inform sample-specific inference.

We formalize this by assuming each observation x_i is drawn from a distribution governed by parameters θ_i:

$$x_i \sim P(x; \theta_i).$$

In population models, the assumption is that $\theta_i = \theta$ for all i. In context-adaptive models, we instead posit that the parameters vary with context:

$$\theta_i = f(c_i) \quad \text{or} \quad \theta_i \sim P(\theta \mid c_i),$$

where c_i captures the relevant covariates or environment for observation i. The goal is to estimate either a deterministic function f or a conditional distribution over parameters.

This shift raises new modeling challenges. Estimating a unique θ_i from a single observation is ill-posed unless we impose structure—smoothness, sparsity, shared representations, or latent grouping. And as adaptivity becomes more implicit (e.g., via neural networks or black-box inference), we must develop tools to recover, interpret, or constrain the underlying parameter variation.

In this review, we examine methods that use context to guide inference, either by specifying how parameters change with covariates or by learning to adapt behavior implicitly. We begin with classical models that impose explicit structure—such as varying-coefficient models and multi-task learning—and then turn to more flexible approaches like meta-learning and in-context learning with foundation models. Though these methods arise from different traditions, they share a common goal: to tailor inference to the local characteristics of each observation or task. Along the way, we highlight recurring themes: complex models often decompose into simpler, context-specific components; foundation models can both adapt to and generate context; and context-awareness challenges classical

assumptions of homogeneity. These perspectives offer a unifying lens on recent advances and open new directions for building adaptive, interpretable, and personalized models.

From Population Assumptions to Context-Adaptive Inference

Most statistical and machine learning models begin with a foundational assumption: that all samples are drawn independently and identically from a shared population distribution. This assumption simplifies estimation and enables generalization from limited data, but it collapses in the presence of meaningful heterogeneity.

In practice, data often reflect differences across individuals, environments, or conditions. These differences may stem from biological variation, temporal drift, site effects, or shifts in measurement context. Treating heterogeneous data as if it were homogeneous can obscure real effects, inflate variance, and lead to brittle predictions.

Failure Modes of Population Models

Even when traditional models appear to fit aggregate data well, they may hide systematic failure modes.

Mode Collapse
When one subpopulation is much larger than another, standard models are biased toward the dominant group, underrepresenting the minority group in both fit and predictions.

Outlier Sensitivity
In the parameter-averaging regime, small but extreme groups can disproportionately distort the global model, especially in methods like ordinary least squares.

Phantom Populations
When multiple subpopulations are equally represented, the global model may fit none of them well, instead converging to a solution that represents a non-existent average case.

These behaviors reflect a deeper problem: the assumption of identically distributed samples is not just incorrect, but actively harmful in heterogeneous settings.

Toward Context-Aware Models

To account for heterogeneity, we must relax the assumption of shared parameters and allow the data-generating process to vary across samples. A general formulation assumes each observation is governed by its own latent parameters:

$$x_i \sim P(x; \theta_i),$$

However, estimating N free parameters from N samples is underdetermined. Context-aware approaches resolve this by introducing structure on how parameters vary, often by assuming that θ_i depends on an observed context c_i:

$$\theta_i = f(c_i) \quad \text{or} \quad \theta_i \sim P(\theta \mid c_i).$$

This formulation makes the model estimable, but it raises new challenges. How should f be chosen? How smooth, flexible, or structured should it be? The remainder of this review explores different

answers to this question, and shows how implicit and explicit representations of context can lead to powerful, personalized models.

Early Remedies: Grouped and Distance-Based Models

Before diving into flexible estimators of $f(c)$, we review early modeling strategies that attempt to break away from homogeneity.

Conditional and Clustered Models

One approach is to group observations into C contexts, either by manually defining conditions (e.g. male vs. female) or using unsupervised clustering. Each group is then assigned a distinct parameter vector:

$$\{\hat{\theta}_0, \ldots, \hat{\theta}_C\} = \arg\max_{\theta_0,\ldots,\theta_C} \sum_{c \in \mathcal{C}} \ell(X_c; \theta_c),$$

where $\ell(X; \theta)$ is the log-likelihood of θ on X and c specifies the covariate group that samples are assigned to. This reduces variance but limits granularity. It assumes that all members of a group share the same distribution and fails to capture variation within a group.

Distance-Regularized Estimation

A more flexible alternative assumes that observations with similar contexts should have similar parameters. This is encoded as a regularization penalty that discourages large differences in θ_i for nearby c_i:

$$\{\hat{\theta}_0, \ldots, \hat{\theta}_N\} = \arg\max_{\theta_0,\ldots,\theta_N} \left(\sum_i \ell(x_i; \theta_i) - \sum_{i,j} \frac{\|\theta_i - \theta_j\|}{D(c_i, c_j)} \right),$$

where $D(c_i, c_j)$ is a distance metric between contexts. This approach allows for smoother parameter variation but requires careful choice of D and regularization strength λ to balance bias and variance. The choice of distance metric D and regularization strength λ controls the bias–variance tradeoff.

Parametric Varying-coefficient models

Original paper (based on a smoothing spline function): [1] Markov networks: [2] Linear varying-coefficient models assume that parameters vary linearly with covariates, a much stronger assumption than the classic varying-coefficient model but making a conceptual leap that allows us to define a form for the relationship between the parameters and covariates.

$$\hat{\theta}_0, \ldots, \hat{\theta}_N = \widehat{A} C^T$$
$$\widehat{A} = \arg\max_A \sum_i \ell(x_i; A c_i)$$

TODO: Note that they achieve distance-matching by using a distance metric under Euclidean distance, which is a special case of the distance-regularized estimation above.

Semi-parametric varying-coefficient Models

Original paper: [3] 2-step estimation with RBF kernels: [4]

Classic varying-coefficient models assume that models with similar covariates have similar parameters, or – more formally – that changes in parameters are smooth over the covariate space. This assumption is encoded as a sample weighting, often using a kernel, where the relevance of a sample to a model is equivalent to its kernel similarity over the covariate space.

$$\hat{\theta}_0, \dots, \hat{\theta}_N = \arg\max_{\theta_0,\dots,\theta_N} \sum_{i,j} \frac{K(c_i, c_j)}{\sum_k K(c_i, c_k)} \ell(x_j; \theta_i)$$

This estimator is the simplest to recover N unique parameter estimates. However, the assumption here is contradictory to the partition model estimator. When the relationship between covariates and parameters is discontinuous or abrupt, this estimator will fail.

Contextualized Models

Seminal work [5] Contextualized ML generalization and applications: [6], [7], [8], [9], [10], [11], [12], [13]

Contextualized models make the assumption that parameters are some function of context, but make no assumption on the form of that function. In this regime, we seek to estimate the function often using a deep learner (if we have some differentiable proxy for probability):

$$\widehat{f} = \arg\max_{f \in \mathcal{F}} \sum_i \ell(x_i; f(c_i))$$

Latent-structure Models

Partition Models

Markov networks: [14] Partition models also assume that parameters can be partitioned into homogeneous groups over the covariate space, but make no assumption about where these partitions occur. This allows the use of information from different groups in estimating a model for a each covariate. Partition model estimators are most often utilized to infer abrupt model changes over time and take the form

$$\hat{\theta}_0, \dots, \hat{\theta}_N = \arg\max_{\theta_0,\dots,\theta_N} \sum_i \ell(x_i; \theta_i) + \sum_{i=2}^{N} \mathrm{TV}(\theta_i, \theta_{i-1})$$

Where the regularizaiton term might take the form

$$\mathrm{TV}(\theta_i, \theta_{i-1}) = |\theta_i - \theta_{i-1}|$$

This still fails to recover a unique parameter estimate for each sample, but gets closer to the spirit of personalized modeling by putting the model likelihood and partition regularizer in competition to find the optimal partitions.

Fine-tuned Models and Transfer Learning

Review: [15] Noted in foundational literature for linear varying coefficient models [3]

Estimate a population model, freeze these parameters, and then include a smaller set of personalized parameters to estimate on a smaller subpopulation.

$$\widehat{\gamma} = \arg\max_{\gamma} = \ell(\gamma; X)$$

$$\widehat{\theta_c} = \arg\max_{\theta_c} = \ell(\theta_c; \widehat{\gamma}, X_c)$$

Context-informed and Latent-structure models

Seminal paper: [16]

Key idea: negative information sharing. Different models should be pushed apart.

$$\hat{\theta}_0, \ldots, \hat{\theta}_N = \arg\max_{\theta_0,\ldots,\theta_N, D} \sum_{i=0}^{N} \prod_{j=0 s.t. D(c_i,c_j)<d}^{N} P(x_j;\theta_i) P(\theta_i;\theta_j)$$

A Spectrum of Context-Awareness

Context-aware models can be viewed along a spectrum of assumptions about the relationship between context and parameters.

Global models: $\theta_i = \theta$ for all i
Grouped models: $\theta_i = \theta_c$ for some finite set of groups
Smooth models: $\theta_i = f(c_i)$, with f assumed to be continuous or low-complexity
Latent models: $\theta_i \sim P(\theta|c_i)$, with f learned implicitly

Each of these choices encodes different beliefs about how parameters vary. The next section formalizes this variation and examines general principles for adaptivity in statistical modeling.

Relevant references:

- Can Subpopulation Shifts Explain Disagreement in Model Generalization? [17]

Principles of Context-Adaptive Inference

What makes a model adaptive? When is it good for a model to be adaptive? While the appeal of adaptivity lies in flexibility and personalized inference, not all adaptivity is good adaptivity. In this section, we formalize the core principles that underlie adaptive modeling.

1. Adaptivity requires flexibility

A model cannot adapt unless it has the capacity to represent multiple behaviors. Flexibility may take the form of nonlinearity, hierarchical structure, or modular components that allow different responses in different settings.

- Interaction effects in regression models [18]
- Hierarchical models that allow for varying effects across groups
- Meta-learning and mixtures-of-experts models that learn to adapt based on context
- Varying-coefficient models that allow coefficients to change with context [1]

2. Adaptivity requires a signal of heterogeneity

- Varying-coefficient models adapt parameters based on observed context [1]
- Contextual bandits adapt actions to context features [19]
- Multi-domain models adapt across known environments or inferred partitions [20]

3. Modularity improves adaptivity

Adaptive systems are easier to design, debug, and interpret when built from modular parts. Modularity supports targeted adaptation, transferability, and disentanglement.

- []

4. Adaptivity implies selectivity

Adaptation must be earned. Overreacting to limited data leads to overfitting. The best adaptive methods include mechanisms for deciding when not to adapt. - Lepski's method [21] - Aggregation of classifiers [22]

5. Adaptivity is bounded by data efficiency

[23]

6. Adaptivity is not a free lunch

Adaptivity improves performance when heterogeneity is real and informative, but it can degrade performance when variation is spurious. Key tradeoffs include:

- **Bias vs. variance**: More flexible adaptation can reduce bias but increase variance
- **Stability vs. personalization**: Highly adaptive models may overfit to noise or adversarial context
- **Inference cost**: Adaptive inference may be more computationally intensive than global prediction

Understanding these tradeoffs is essential when designing systems for real-world deployment.

When Adaptivity Fails: Common Failure Modes

Even when all the ingredients are present, adaptivity can backfire. Common failure modes include:

- Spurious Adaptation: Adapting to unstable or confounded features [24]
- Overfitting in Low-Data Contexts: Attempting fine-grained adaptation with insufficient signal
- Modularity Mis-specification: Adapting in the wrong units or groupings [25]
- Feedback Loops: Models that change the data distribution they rely on [26]

Related references:

Explicit Adaptivity: Structured Estimation of $f(c)$

In classical statistical modeling, all observations are typically assumed to share a common set of parameters. However, modern datasets often display significant heterogeneity across individuals, locations, or experimental conditions, making this assumption unrealistic in many real-world applications. To better capture such heterogeneity, recent approaches model parameters as explicit functions of observed context, formalized as $\theta_i = f(c_i)$, where f maps each context to a sample-specific parameter [27].

This section systematically reviews explicit adaptivity methods, with a focus on structured estimation of $f(c)$. We begin by revisiting classical varying-coefficient models, which provide a conceptual and methodological foundation for modeling context-dependent effects. We then categorize recent advances in explicit adaptivity according to three principal strategies for estimating $f(c)$: (1) smooth nonparametric models that generalize classical techniques, (2) structurally constrained models that incorporate domain-specific knowledge such as spatial or network structure, and (3) learned function approximators that leverage machine learning methods for high-dimensional or complex contexts.

Finally, we summarize key theoretical developments and highlight promising directions for future research in this rapidly evolving field.

Classical Varying-Coefficient Models: A Foundation

Varying-coefficient models (VCMs) are a foundational tool for modeling heterogeneity, as they allow model parameters to vary smoothly with observed context variables [27,28,29]. In their original formulation, the regression coefficients are treated as nonparametric functions of low-dimensional covariates, such as time or age. The standard VCM takes the form

$$y_i = \sum_{j=1}^{p} \beta_j(c_i) x_{ij} + \varepsilon_i,$$

where each $\beta_j(c)$ is an unknown smooth function, typically estimated using kernel smoothing, local polynomials, or penalized splines [28].

This approach provides greater flexibility than fixed-coefficient models and is widely used for longitudinal and functional data analysis. The assumption of smoothness makes estimation and theoretical analysis more tractable, but also imposes limitations. Classical VCMs work best when the context is low-dimensional and continuous. They may struggle with abrupt changes, discontinuities, or high-dimensional and structured covariates. In such cases, interpretability and accuracy can be compromised, motivating the development of a variety of modern extensions, which will be discussed in the following sections.

Advances in Modeling $f(c)$

Recent years have seen substantial progress in the modeling of $f(c)$, the function mapping context to model parameters. These advances can be grouped into three major strategies: (1) **smooth non-parametric models** that extend classical flexibility; (2) **structurally constrained approaches** that encode domain knowledge such as spatial or network topology; and (3) high-capacity **learned function approximators** from machine learning designed for high-dimensional, unstructured contexts. Each strategy addresses specific challenges in modeling heterogeneity, and together they provide a comprehensive toolkit for explicit adaptivity.

Smooth Non-parametric Models

This family of models generalizes the classical VCM by expressing $f(c)$ as a flexible, smooth function estimated with basis expansions and regularization. Common approaches include spline-based methods, local polynomial regression, and RKHS-based frameworks. For instance, [28] developed a semi-nonparametric VCM using RKHS techniques for imaging genetics, enabling the model to capture complex nonlinear effects. Such methods are central to generalized additive models, supporting both flexibility and interpretability. Theoretical work has shown that penalized splines and kernel methods offer strong statistical guarantees in moderate dimensions, although computational cost and overfitting can become issues as the dimension of c increases.

Structurally Constrained Models

Another direction focuses on incorporating structural information into $f(c)$, especially when the context is discrete, clustered, or topologically organized.

Piecewise-Constant and Partition-Based Models. Here, model parameters are allowed to remain constant within specific regions or clusters of the context space, rather than vary smoothly.

Approaches include classical grouped estimators and modern partition models, which may learn changepoints using regularization tools like total variation penalties or the fused lasso. This framework is particularly effective for data with abrupt transitions or heterogeneous subgroups.

Structured Regularization for Spatial, Graph, and Network Data. When context exhibits known structure, regularization terms can be designed to promote similarity among neighboring coefficients [30]. For example, spatially varying-coefficient models have been applied to problems in geographical analysis and econometrics, where local effects are expected to vary across adjacent regions [31,32,33,34]. On networked data, the network VCM of [35] generalizes these ideas by learning both the latent positions and the parameter functions on graphs, allowing the model to accommodate complex relational heterogeneity. Such structural constraints allow models to leverage domain knowledge, improving efficiency and interpretability where smooth models may struggle.

Learned Function Approximators

A third class of methods is rooted in modern machine learning, leveraging high-capacity models to approximate $f(c)$ directly from data. These approaches are especially valuable when the context is high-dimensional or unstructured, where classical assumptions may no longer be sufficient.

Tree-Based Ensembles. Gradient boosting decision trees (GBDTs) and related ensemble methods are well suited to tabular and mixed-type contexts. The framework developed by [36] extends varying-coefficient models by integrating gradient boosting, achieving strong predictive performance with a level of interpretability. These models are typically easier to train and tune than deep neural networks, and their structure lends itself to interpretation with tools such as SHAP.

Deep Neural Networks. For contexts defined by complex, high-dimensional features such as images, text, or sequential data, deep neural networks offer unique advantages for modeling $f(c)$. These architectures can learn adaptive, data-driven representations that capture intricate relationships beyond the scope of classical models. Applications include personalized medicine, natural language processing, and behavioral science, where outcomes may depend on subtle or latent features of the context.

The decision between these machine learning approaches depends on the specific characteristics of the data, the priority placed on interpretability, and computational considerations. Collectively, these advances have significantly broadened the scope of explicit adaptivity, making it feasible to model heterogeneity in ever more complex settings.

Key Theoretical Advances

The expanding landscape of varying-coefficient models (VCMs) has been supported by substantial theoretical progress, which secures the validity of flexible modeling strategies and guides their practical use. The nature of these theoretical results often reflects the core structural assumptions of each model class.

Theory for Smooth Non-parametric Models. For classical VCMs based on kernel smoothing, local polynomial estimation, or penalized splines, extensive theoretical work has characterized their convergence rates and statistical efficiency. Under standard regularity conditions, these estimators are known to achieve minimax optimality for function estimation in moderate dimensions [27]. Recent developments, such as the work of [28], have established asymptotic normality in semi-nonparametric settings, which enables valid confidence interval construction and hypothesis testing even in complex applications.

Theory for Structurally Constrained Models. When discrete or network structure is incorporated into VCMs, theoretical analysis focuses on identifiability, regularization properties, and conditions for consistent estimation. For example, [35] provide non-asymptotic error bounds for estimators in network VCMs, demonstrating that consistency can be attained when the underlying graph topology satisfies certain connectivity properties. In piecewise-constant and partition-based models, results from change-point analysis and total variation regularization guarantee that abrupt parameter changes can be recovered accurately under suitable sparsity and signal strength conditions.

Theory for High-Capacity and Learned Models. The incorporation of machine learning models into VCMs introduces new theoretical challenges. For high-dimensional and sparse settings, oracle inequalities and penalized likelihood theory establish conditions for consistent variable selection and accurate estimation, as seen in methods based on boosting and other regularization techniques [36,37]. In the context of neural network-based VCMs, the theory is still developing, with current research focused on understanding generalization properties and identifiability in non-convex optimization. This remains an active and important frontier for both statistical and machine learning communities.

These theoretical advances provide a rigorous foundation for explicit adaptivity, ensuring that VCMs can be deployed confidently across a wide range of complex and structured modeling scenarios.

Synthesis and Future Directions

Selecting an appropriate modeling strategy for $f(c)$ involves weighing flexibility, interpretability, computational cost, and the extent of available domain knowledge. Learned function approximators, such as deep neural networks, offer unmatched capacity for modeling complex, high-dimensional relationships. However, classical smooth models and structurally constrained approaches often provide greater interpretability, transparency, and statistical efficiency. The choice of prior assumptions and the scalability of the estimation procedure are also central considerations in applied contexts.

Looking forward, several trends are shaping the field. One important direction is the integration of varying-coefficient models with foundation models from natural language processing and computer vision. By using pre-trained embeddings as context variables c_i, it becomes possible to incorporate large amounts of prior knowledge and extend VCMs to multi-modal and unstructured data sources. Another active area concerns the principled combination of cross-modal contexts, bringing together information from text, images, and structured covariates within a unified VCM framework.

Advances in interpretability and visualization for high-dimensional or black-box coefficient functions are equally important. Developing tools that allow users to understand and trust model outputs is critical for the adoption of VCMs in sensitive areas such as healthcare and policy analysis.

Finally, closing the gap between methodological innovation and practical deployment remains a priority. Although the literature has produced many powerful variants of VCMs, practical adoption is often limited by the availability of software and the clarity of methodological guidance [29]. Continued investment in user-friendly implementations, open-source libraries, and empirical benchmarks will facilitate broader adoption and greater impact.

In summary, explicit adaptivity through structured estimation of $f(c)$ now forms a core paradigm at the interface of statistical modeling and machine learning. Future progress will focus not only on expanding the expressive power of these models, but also on making them more accessible, interpretable, and practically useful in real-world applications.

Implicit Adaptivity: Emergent Contextualization within Complex Models

Not all models adapt through explicit parameterization. In many modern systems, adaptation emerges from architecture, training data, or inference dynamics—without being hard-coded as a function of context.

We refer to this as *implicit adaptivity*. These methods do not model θ_i directly as a function of c_i, nor do they always define context formally. Instead, they internalize patterns across training distributions in a way that enables flexible behavior at inference time.

A canonical example is **in-context learning** with foundation models. Given a prompt consisting of a few examples, the model adjusts its behavior—often achieving personalization or task adaptation—without updating weights or making any explicit inference over θ. This capacity arises from pretraining on diverse data and from the model's architecture, not from structured estimation.

Other forms of implicit adaptivity include:

- **Fine-tuned models** that generalize across tasks or domains by adjusting shared components.
- **Attention-based architectures** that condition on context without defining a parametric mapping.
- **Gradient-based meta-learners** trained to produce fast adaptation without modeling $\theta(c)$ explicitly.

These methods challenge the boundary between training and inference. They blur the distinction between model parameters and data inputs, and they rely on massive-scale training to amortize the cost of adaptation.

In this section, we examine:

- How implicit adaptivity arises in foundation models
- What assumptions these models make (implicitly or explicitly) about context
- How their performance compares to structured, explicit approaches
- When it's valuable to make the adaptation process more interpretable or modular

Implicit adaptivity offers powerful capabilities, but it also hides structure that could be useful for analysis, debugging, or control. The next section explores efforts to *make the implicit explicit*—by approximating, interpreting, or extracting the latent adaptation mechanisms inside black-box models.

Defining Implicit Adaptation

Neural Networks with context inputs (e.g. interaction effects, attention mechanisms, etc.)

Amortized Inference and Meta-Learning

In-context learning in transformers and foundation models

Making Implicit Adaptivity Explicit: Local Models, Surrogates and Post Hoc Approximations

This section focuses on methods that aim to extract, approximate, or control the internal adaptivity mechanisms of black-box models. These approaches recognize that implicit adaptivity—while powerful—can be opaque, hard to debug, and brittle to distribution shift. By surfacing structure, we gain interpretability, composability, and sometimes improved generalization.

Motivation

- Implicit adaptivity can succeed without explicit modeling, but:
  - It obscures *why* and *how* a model adapts
  - It limits modular reuse and inspection
  - It makes personalization hard to constrain or audit
- Making adaptivity explicit supports:
  - Better alignment with downstream goals
  - Composability of learned modules
  - Debugging and error attribution

Approaches

Surrogate Modeling

- Fit interpretable surrogates (e.g., linear models, decision trees) to approximate model behavior locally
- Applications:
  - Explaining predictions post-hoc
  - Approximating $f(c)$ from input-output behavior
- References:
  - LIME, SHAP

Prototype and Nearest-Neighbor Methods

- Use nearest neighbors in representation space to approximate model adaptation
- Enables interpretability and modular updates
- Related to contextual bandits, exemplar models

Amortization Diagnostics

- For amortized inference (e.g., variational autoencoders), analyze encoder mappings to understand how $q(\theta|x)$ varies with x
- Could treat encoder as a learned $f(c)$ and evaluate its fidelity

Disentangled Representations

- Train models with constraints (e.g., variational regularization, info bottlenecks) to encourage explicit factors of variation
- Goal: make parameter changes traceable to distinct contextual causes

Parameter Extraction

- Techniques like linear probes, weight attribution, or synthetic tasks to reverse-engineer how models adapt internally
- Example: “what part of the weights encode the task?”

Tradeoffs

- Fidelity vs interpretability
- Local vs global explanations
- Approximation error vs modular control

Open Questions

- Can we extract *portable* modules from foundation models?
- When does making structure explicit improve performance?
- What is the right level of abstraction—parameters, functions, latent causes?

This section bridges black-box adaptation and structured inference. It highlights how interpretability and performance need not be at odds—especially when the goal is robust, composable, and trustworthy adaptation.

TODO: Discussing the implications of context-adaptive interpretations for traditional models. Related work including LIME/DeepLift/DeepSHAP.

Relevant references:

- [38]
- Interpretations are statistics [39]

Context-Invariant Training: A View from the Converse

TODO: The converse of context-adaptive models, exploring the implications of training context-invariant models. e.g. out-of-distribution generalization, robustness to adversarial attacks.

Relevant references:

- Invariant Risk Minimization [40]
- Out-of-Distribution Generalization via Risk Extrapolation [41]
- The Risks of Invariant Risk Minimization [24]
- Conditional Variance Penalties and Domain Adaptation [42]
- Can Subpopulation Shifts Explain Disagreement in Model Generalization? [17]

Adversarial Robustness as Context-Invariant Training

Related references:

- Towards Deep Learning Models Resistant to Adversarial Attacks [43]
- Robustness May Be at Odds with Accuracy [44]

Training methods for Context-Invariant Models

- Just Train Twice: Improving Group Robustness without Training Group Information [45]
- Environment Inference for Invariant Learning [46]
- Distributionally Robust Neural Networks for Group Shifts [25]

Applications, Case Studies, Evaluation Metrics, and Tools

Implementation Across Sectors

TODO: Detailed examination of context-adaptive models in sectors like healthcare and finance.

Relevant references:

- [47]
- [48]

Performance Evaluation

TODO: Successes, failures, and comparative analyses of context-adaptive models across applications.

Survey of Tools

TODO: Reviewing current technological supports for context-adaptive models.

Selection and Usage Guidance

TODO: Offering practical advice on tool selection and use for optimal outcomes.

Future Trends and Opportunities with Foundation Models

Emerging Technologies

TODO: Identifying upcoming technologies and predicting their impact on context-adaptive learning.

Advances in Methodologies

TODO: Speculating on potential future methodological enhancements.

Expanding Frameworks with Foundation Models

Foundation models refer to large-scale, general-purpose neural networks, predominantly transformer-based architectures, trained on vast datasets using self-supervised learning [49]. These models have significantly transformed modern statistical modeling and machine learning due to their flexibility, adaptability, and strong performance across diverse domains. Notably, large language models (LLMs) such as GPT-4 [50] and LLaMA-3.1 [51] have achieved substantial advancements in natural language processing (NLP), demonstrating proficiency in tasks ranging from text generation and summarization to question-answering and dialogue systems. Beyond NLP, foundation models also excel in multimodal (text-vision) tasks [52], text embedding generation [53], and structured tabular data analysis [54], highlighting their broad applicability.

A key strength of foundation models lies in their capacity to dynamically adapt to different contexts provided by inputs. This adaptability is primarily achieved through techniques such as prompting, which involves designing queries to guide the model's behavior implicitly, allowing task-specific responses without additional fine-tuning [55]. Furthermore, mixture-of-experts (MoE) architectures amplify this contextual adaptability by employing routing mechanisms that select specialized sub-models or "experts" tailored to specific input data, thus optimizing computational efficiency and performance [56].

Foundation Models as Context

Foundation models offer significant opportunities by supplying context-aware information that enhances various stages of statistical modeling and inference:

Feature Extraction and Interpretation: Foundation models transform raw, unstructured data into structured and interpretable representations. For example, targeted prompts enable LLMs to extract insightful features from text, providing meaningful insights and facilitating interpretability [59]. This allows statistical models to operate directly on semantically meaningful features rather than on raw, less interpretable data.

Contextualized Representations for Downstream Modeling: Foundation models produce adaptable embeddings and intermediate representations useful as inputs for downstream models, such as decision trees or linear models [60]. These embeddings significantly enhance the training of both complex, black-box models [61] and simpler statistical methods like n-gram-based analyses [62], thereby broadening the application scope and effectiveness of statistical approaches.

Post-hoc Interpretability: Foundation models support interpretability by generating natural-language explanations for decisions made by complex models. This capability enhances transparency and trust in statistical inference, providing clear insights into how and why certain predictions or decisions are made [63].

Recent innovations underscore the role of foundation models in context-sensitive inference and enhanced interpretability:

FLAN-MoE (Fine-tuned Language Model with Mixture of Experts) [64] combines instruction tuning with expert selection, dynamically activating relevant sub-models based on the context. This method significantly improves performance across diverse NLP tasks, offering superior few-shot and zero-shot capabilities. It also facilitates interpretability through explicit expert activations. Future directions may explore advanced expert-selection techniques and multilingual capabilities.

LMPriors (Pre-Trained Language Models as Task-Specific Priors) [65] leverages semantic insights from pre-trained models like GPT-3 to guide tasks such as causal inference, feature selection, and reinforcement learning. This method markedly enhances decision accuracy and efficiency without requiring extensive supervised datasets. However, it necessitates careful prompt engineering to mitigate biases and ethical concerns.

Mixture of In-Context Experts (MoICE) [65] introduces a dynamic routing mechanism within attention heads, utilizing multiple Rotary Position Embeddings (RoPE) angles to effectively capture token positions in sequences. MoICE significantly enhances performance on long-context sequences and retrieval-augmented generation tasks by ensuring complete contextual coverage. Efficiency is achieved through selective router training, and interpretability is improved by explicitly visualizing attention distributions, providing detailed insights into the model's reasoning process.

Open Problems

Theoretical Challenges

TODO: Critically examining unresolved theoretical issues like identifiability, etc.

Ethical and Regulatory Considerations

TODO: Discussing the ethical landscape and regulatory challenges, with focus on benefits of interpretability and regulatability.

Complexity in Implementation

TODO: Addressing obstacles in practical applications and gathering insights from real-world data.

TODO: Other open problems?

Conclusion

Overview of Insights

TODO: Summarizing the main findings and contributions of this review.

Future Directions

TODO: Discussing potential developments and innovations in context-adaptive statistical inference.

References

1. **Varying-Coefficient Models**
Trevor Hastie, Robert Tibshirani
Journal of the Royal Statistical Society Series B: Statistical Methodology (1993-09-01)
https://doi.org/gmfvmb
DOI: 10.1111/j.2517-6161.1993.tb01939.x

2. **Bayesian Edge Regression in Undirected Graphical Models to Characterize Interpatient Heterogeneity in Cancer**
Zeya Wang, Veerabhadran Baladandayuthapani, Ahmed O Kaseb, Hesham M Amin, Manal M Hassan, Wenyi Wang, Jeffrey S Morris
Journal of the American Statistical Association (2022-01-05) https://doi.org/gt68hr
DOI: 10.1080/01621459.2021.2000866 · PMID: 36090952 · PMCID: PMC9454401

3. **Statistical estimation in varying coefficient models**
Jianqing Fan, Wenyang Zhang
The Annals of Statistics (1999-10-01) https://doi.org/dsxd4s
DOI: 10.1214/aos/1017939139

4. **Time-Varying Coefficient Model Estimation Through Radial Basis Functions**
Juan Sosa, Lina Buitrago
arXiv (2021-03-02) https://arxiv.org/abs/2103.00315

5. **Contextual Explanation Networks**
Maruan Al-Shedivat, Avinava Dubey, Eric P Xing
arXiv (2017) https://doi.org/gt68h9
DOI: 10.48550/arxiv.1705.10301

6. **Contextualized Machine Learning**
Benjamin Lengerich, Caleb N Ellington, Andrea Rubbi, Manolis Kellis, Eric P Xing
arXiv (2023) https://doi.org/gt68jg
DOI: 10.48550/arxiv.2310.11340

7. **NOTMAD: Estimating Bayesian Networks with Sample-Specific Structures and Parameters**
Ben Lengerich, Caleb Ellington, Bryon Aragam, Eric P Xing, Manolis Kellis
arXiv (2021) https://doi.org/gt68jc
DOI: 10.48550/arxiv.2111.01104

8. **Contextualized: Heterogeneous Modeling Toolbox**
Caleb N Ellington, Benjamin J Lengerich, Wesley Lo, Aaron Alvarez, Andrea Rubbi, Manolis Kellis, Eric P Xing
Journal of Open Source Software (2024-05-08) https://doi.org/gt68h8
DOI: 10.21105/joss.06469

9. **Contextualized Policy Recovery: Modeling and Interpreting Medical Decisions with Adaptive Imitation Learning**
Jannik Deuschel, Caleb N Ellington, Yingtao Luo, Benjamin J Lengerich, Pascal Friederich, Eric P Xing
arXiv (2023) https://doi.org/gt68jf
DOI: 10.48550/arxiv.2310.07918

10. **Automated interpretable discovery of heterogeneous treatment effectiveness: A COVID-19 case study**
Benjamin J Lengerich, Mark E Nunnally, Yin Aphinyanaphongs, Caleb Ellington, Rich Caruana
Journal of Biomedical Informatics (2022-06) https://doi.org/gt68h5
DOI: 10.1016/j.jbi.2022.104086 · PMID: 35504543 · PMCID: PMC9055753

11. **Discriminative Subtyping of Lung Cancers from Histopathology Images via Contextual Deep Learning**
Benjamin J Lengerich, Maruan Al-Shedivat, Amir Alavi, Jennifer Williams, Sami Labbaki, Eric P Xing
Cold Spring Harbor Laboratory (2020-06-26) https://doi.org/gt68h6
DOI: 10.1101/2020.06.25.20140053

12. **Learning to Estimate Sample-specific Transcriptional Networks for 7000 Tumors**
Caleb N Ellington, Benjamin J Lengerich, Thomas BK Watkins, Jiekun Yang, Abhinav Adduri, Sazan Mahbub, Hanxi Xiao, Manolis Kellis, Eric P Xing
Cold Spring Harbor Laboratory (2023-12-04) https://doi.org/gt68h7
DOI: 10.1101/2023.12.01.569658

13. **Contextual Feature Selection with Conditional Stochastic Gates**
Ram Dyuthi Sristi, Ofir Lindenbaum, Shira Lifshitz, Maria Lavzin, Jackie Schiller, Gal Mishne, Hadas Benisty
arXiv (2023) https://doi.org/gt68jh
DOI: 10.48550/arxiv.2312.14254

14. **Estimating time-varying networks**
Mladen Kolar, Le Song, Amr Ahmed, Eric P Xing
The Annals of Applied Statistics (2010-03-01) https://doi.org/b3rn6q
DOI: 10.1214/09-aoas308

15. **When Personalization Harms: Reconsidering the Use of Group Attributes in Prediction**
Vinith M Suriyakumar, Marzyeh Ghassemi, Berk Ustun
arXiv (2022) https://doi.org/gt68jd
DOI: 10.48550/arxiv.2206.02058

16. **Learning Sample-Specific Models with Low-Rank Personalized Regression**
Benjamin Lengerich, Bryon Aragam, Eric P Xing
arXiv (2019) https://doi.org/gt68jb
DOI: 10.48550/arxiv.1910.06939

17. **Sketch-Based Anomaly Detection in Streaming Graphs**
Siddharth Bhatia, Mohit Wadhwa, Kenji Kawaguchi, Neil Shah, Philip S Yu, Bryan Hooi
arXiv (2023-07-18) https://arxiv.org/abs/2106.04486

18. **Intelligible Models for HealthCare**
Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, Noemie Elhadad
Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2015-08-10) https://doi.org/gftgxk
DOI: 10.1145/2783258.2788613

19. **Adapting multi-armed bandits policies to contextual bandits scenarios**
David Cortes
arXiv (2019-11-26) https://arxiv.org/abs/1811.04383

20. **Environment Inference for Invariant Learning**
Elliot Creager, JÃ¶rn-Henrik Jacobsen, Richard Zemel

arXiv (2021-07-16) https://arxiv.org/abs/2010.07249

21. **Lepski's Method and Adaptive Estimation of Nonlinear Integral Functionals of Density**
Rajarshi Mukherjee, Eric Tchetgen Tchetgen, James Robins
arXiv (2016-01-12) https://arxiv.org/abs/1508.00249

22. **Optimal Rates of Aggregation**
Alexandre B Tsybakov
Lecture Notes in Computer Science (2003) https://doi.org/czntw5
DOI: 10.1007/978-3-540-45167-9_23

23. **Optimal Estimation of Change in a Population of Parameters**
Ramya Korlakai Vinayak, Weihao Kong, Sham M Kakade
arXiv (2019-12-02) https://arxiv.org/abs/1911.12568

24. **The Risks of Invariant Risk Minimization**
Elan Rosenfeld, Pradeep Ravikumar, Andrej Risteski
arXiv (2021-03-30) https://arxiv.org/abs/2010.05761

25. **Distributionally Robust Neural Networks for Group Shifts: On the Importance of Regularization for Worst-Case Generalization**
Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, Percy Liang
arXiv (2020-04-03) https://arxiv.org/abs/1911.08731

26. **The Selective Labels Problem**
Himabindu Lakkaraju, Jon Kleinberg, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan
Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2017-08-04) https://doi.org/ggd7hz
DOI: 10.1145/3097983.3098066 · PMID: 29780658 · PMCID: PMC5958915

27. **Varying-coefficient models**
Trevor Hastie, Robert Tibshirani
Journal of the Royal Statistical Society: Series B (Methodological)

28. **Semi-nonparametric Varying Coefficients Models for Imaging Genetics**
Ting Li, Yang Yu, Xiao Wang, JS Marron, Hongtu Zhu
Statistica Sinica
DOI: 10.5705/ss.202024.0118

29. **Publication Trends on the Varying Coefficients Model: Estimating the Actual (Under)Utilization of a Highly Acclaimed Method for Studying Statistical Interactions**
Assaf Botzer
Publications
DOI: 10.3390/publications13020019

30. **Graph-Regularized Estimation for Context-Varying Models**
Yu Shi, Yang Liu, Hao Yan

31. **Fast Spatio-Temporally Varying Coefficient Modeling With Reluctant Interaction Selection**
Daisuke Murakami, Shinichiro Shirota, Seiji Kajita, Mami Kajita
Geographical Analysis
DOI: 10.1111/gean.70005

32. **Spatially Varying Coefficient Models for Estimating Heterogeneous Mixture Effects**
Jacob Englert, Howard Chang

33. **Varying-Coefficient Panel Models with Spatial Dependence**
Yiqing Hu, Qingyuan Zhao
Journal of Econometrics
DOI: 10.1016/j.jeconom.2024.105883

34. **Urban Economic Modeling with Spatially Varying Coefficients**
Chongliang Luo, Yihong Du, Peng Zhao
Regional Science and Urban Economics
DOI: 10.1016/j.regsciurbeco.2024.104009

35. **Network Varying Coefficient Model**
Xinyan Fan, Kuangnan Fang, Wei Lan, Chih-Ling Tsai
Journal of the American Statistical Association
DOI: 10.1080/01621459.2025.2470481

36. **Boosted Trees for Varying-Coefficient Models**
Yunfei Wang, Qiang Sun
Machine Learning
DOI: 10.1007/s00180-025-01603-8

37. **XGBoost-Inspired Estimation for High-Dimensional Varying Coefficient Models**
Yu Cheng, Dongdong Yang, Denny Zhou

38. **In-Context Explainers: Harnessing LLMs for Explaining Black Box Models**
Nicholas Kroeger, Dan Ley, Satyapriya Krishna, Chirag Agarwal, Himabindu Lakkaraju
arXiv (2024-07-12) https://arxiv.org/abs/2310.05797

39. **Rethinking Explainable Machine Learning as Applied Statistics**
Sebastian Bordt, Eric Raidl, Ulrike von Luxburg
arXiv (2025-06-17) https://arxiv.org/abs/2402.02870

40. **Invariant Risk Minimization**
Martin Arjovsky, LÃ©on Bottou, Ishaan Gulrajani, David Lopez-Paz
arXiv (2020-03-31) https://arxiv.org/abs/1907.02893

41. **Out-of-Distribution Generalization via Risk Extrapolation (REx)**
David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Remi Le Priol, Aaron Courville
arXiv (2021-02-26) https://arxiv.org/abs/2003.00688

42. **Conditional Variance Penalties and Domain Shift Robustness**
Christina Heinze-Deml, Nicolai Meinshausen
arXiv (2019-04-16) https://arxiv.org/abs/1710.11469

43. **Towards Deep Learning Models Resistant to Adversarial Attacks**
Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, Adrian Vladu
arXiv (2019-09-06) https://arxiv.org/abs/1706.06083

44. **Robustness May Be at Odds with Accuracy**
Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, Aleksander Madry
arXiv (2019-09-10) https://arxiv.org/abs/1805.12152

45. **On the Sample Complexity of Adversarial Multi-Source PAC Learning**
Nikola Konstantinov, Elias Frantar, Dan Alistarh, Christoph H Lampert
arXiv (2020-07-01) https://arxiv.org/abs/2002.10384

46. **Conflict-Averse Gradient Descent for Multi-task Learning**
Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, Qiang Liu
arXiv (2024-02-22) https://arxiv.org/abs/2110.14048

47. **Exact Inference for Transformed Large-Scale Varying Coefficient Models with Applications**
Tianyu Chen, Robert Habans, Thomas Douthat, Jenna Losh, Lida Chalangar Jalili Dehkharghani, Li-Hsiang Lin
Journal of Data Science (2025-01-01) https://doi.org/g9t2rs
DOI: 10.6339/25-jds1181

48. **Variable Selection for Generalized Single-Index Varying-Coefficient Models with Applications to Synergistic G × E Interactions**
Shunjie Guan, Xu Liu, Yuehua Cui
Mathematics (2025-01-31) https://doi.org/g9t2rp
DOI: 10.3390/math13030469

49. **On the Opportunities and Risks of Foundation Models**
Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, … Percy Liang
arXiv (2021) https://doi.org/hw3v
DOI: 10.48550/arxiv.2108.07258

50. **GPT-4 Technical Report**
OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, … Barret Zoph
arXiv (2023) https://doi.org/grx4cb
DOI: 10.48550/arxiv.2303.08774

51. **The Llama 3 Herd of Models**
Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, … Zhiyu Ma
arXiv (2024) https://doi.org/ndw6
DOI: 10.48550/arxiv.2407.21783

52. **Learning Transferable Visual Models From Natural Language Supervision**
Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, … Ilya Sutskever
arXiv (2021) https://doi.org/hs7z
DOI: 10.48550/arxiv.2103.00020

53. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**
Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova
arXiv (2018) https://doi.org/hm65
DOI: 10.48550/arxiv.1810.04805

54. **TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second**
Noah Hollmann, Samuel Müller, Katharina Eggensperger, Frank Hutter
arXiv (2022) https://doi.org/g9t22b
DOI: 10.48550/arxiv.2207.01848

55. **Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing**
Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, Graham Neubig
ACM Computing Surveys (2023-01-16) https://doi.org/gq5fh2

DOI: 10.1145/3560815

56. **Mixture of experts: a literature survey**
Saeed Masoudnia, Reza Ebrahimpour
Artificial Intelligence Review (2012-05-12) https://doi.org/f59sxs
DOI: 10.1007/s10462-012-9338-y

57. **CHiLL: Zero-shot Custom Interpretable Feature Extraction from Clinical Notes with Large Language Models**
Denis Jered McInerney, Geoffrey Young, Jan-Willem van de Meent, Byron C Wallace
arXiv (2023) https://doi.org/g9t22g
DOI: 10.48550/arxiv.2302.12343

58. **Learning Interpretable Style Embeddings via Prompting LLMs**
Ajay Patel, Delip Rao, Ansh Kothary, Kathleen McKeown, Chris Callison-Burch
arXiv (2023) https://doi.org/g9t22h
DOI: 10.48550/arxiv.2305.12696

59. **Tree Prompting: Efficient Task Adaptation without Fine-Tuning**
Chandan Singh, John Morris, Alexander Rush, Jianfeng Gao, Yuntian Deng
Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (2023) https://doi.org/gtgrkq
DOI: 10.18653/v1/2023.emnlp-main.384

60. **What Can Transformers Learn In-Context? A Case Study of Simple Function Classes**
Shivam Garg, Dimitris Tsipras, Percy Liang, Gregory Valiant
arXiv (2022) https://doi.org/g9t22c
DOI: 10.48550/arxiv.2208.01066

61. **One Embedder, Any Task: Instruction-Finetuned Text Embeddings**
Hongjin Su, Weijia Shi, Jungo Kasai, Yizhong Wang, Yushi Hu, Mari Ostendorf, Wen-tau Yih, Noah A Smith, Luke Zettlemoyer, Tao Yu
arXiv (2022) https://doi.org/g9t22f
DOI: 10.48550/arxiv.2212.09741

62. **Augmenting interpretable models with large language models during training**
Chandan Singh, Armin Askari, Rich Caruana, Jianfeng Gao
Nature Communications (2023-11-30) https://doi.org/g9t2z9
DOI: 10.1038/s41467-023-43713-1 · PMID: 38036543 · PMCID: PMC10689442

63. **Explaining Datasets in Words: Statistical Models with Natural Language Parameters**
Ruiqi Zhong, Heng Wang, Dan Klein, Jacob Steinhardt
arXiv (2024) https://doi.org/g9t22k
DOI: 10.48550/arxiv.2409.08466

64. **Mixture-of-Experts Meets Instruction Tuning:A Winning Combination for Large Language Models**
Sheng Shen, Le Hou, Yanqi Zhou, Nan Du, Shayne Longpre, Jason Wei, Hyung Won Chung, Barret Zoph, William Fedus, Xinyun Chen, … Denny Zhou
arXiv (2023) https://doi.org/g9t22j
DOI: 10.48550/arxiv.2305.14705

65. **LMPriors: Pre-Trained Language Models as Task-Specific Priors**
Kristy Choi, Chris Cundy, Sanjari Srivastava, Stefano Ermon
arXiv (2022) https://doi.org/g9t22d
DOI: 10.48550/arxiv.2210.12530