Causal Discovery using Model Invariance through Knockoff Interventions

Wasim Ahmad [1] **Maha Shadaydeh** [1] **Joachim Denzler** [1 2]

Abstract

Cause-effect analysis is crucial to understand the underlying mechanism of a system. We propose to exploit model invariance through interventions on the predictors to infer causality in nonlinear multivariate systems of time series. We model nonlinear interactions in time series using DeepAR and then expose the model to different environments using Knockoffs-based interventions to test model invariance. Knockoff samples are pairwise exchangeable, in-distribution and statistically null variables generated without knowing the response. We test model invariance where we show that the distribution of the response residual does not change significantly upon interventions on non-causal predictors. We evaluate our method on real and synthetically generated time series. Overall our method outperforms other widely used causality methods, i.e, VAR Granger causality, VARLiNGAM and PCMCI$^+$.

1. Introduction

Discovering causal graph in a system is vital to understand system's dynamics and its response to changes in environment. The methods which extract causal information from observational data are crucial for the broad field of artificial intelligence. In this work we use deep networks to learn nonlinear interactions in multivariate time series and test model invariance through Knockoffs (Barber & Candès, 2015; Candes et al., 2017) based interventions for causal inference. We build our work on the assumption that the causal mechanism in a real-world system does not change across diverse environments (Peters et al., 2016; Heinze-Deml et al., 2018; Pfister et al., 2019). More specifically, the conditional distribution of the response given its causal predictors remains invariant for all observations over different settings. A model is regarded as invariant if the distribution of the residuals of model's response does not change when intervened on its non-causal predictors.

Deep networks best suit to learn nonlinear interactions in a multivariate system. We use DeepAR (Salinas et al., 2020) to model nonlinear relations in multivariate time series, which is based on recurrent neural networks (RNNs) and generate probabilistic forecast. It characterizes conditioning on observed variables and estimate latent features i.e. trends and seasonality from time series that might help in the identification of spurious associations. To expose trained models to unseen settings for causal inference, we can either omit or replace model predictors with a different representation. Since, deep networks cannot handle missing input or out-of-distribution (OOD) data, we propose to use the theoretically well-established Knockoffs framework to generate interventional environments.

Knockoffs are statistically null-variables, in-distribution and as uncorrelated as possible with the originals. Moreover, Knockoffs possess the pairwise exchangeability property when means that replacing any variable with their knockoff copy in a multivariate setup will not change the underlying joint distribution of the set of variables. We highlight the effectiveness of Knockoffs as interventional variables by comparing it with other intervention methods such as OOD, mean and uniform distribution. We discover full causal graph assuming causal sufficiency and stationarity in synthetic and real data and show that our method outperforms widely used causality methods, i.e., VAR Granger causality (VAR-GC) (Granger, 1969), VARLiNGAM (Hyvärinen et al., 2010), PCMCI$^+$ (Runge, 2020) as well as other applied intervention methods.

The remaining of this paper is organized as follows. In Section 2, we cover the related work. We provide methodological background in Section 3. In Section 4, we describe our proposed method. Results are presented and discussed in Section 5. In Section 6 we conclude our work.

2. Related Work

The methods in (Peters et al., 2016; Heinze-Deml et al., 2018; Pfister et al., 2019; Gamella & Heinze-Deml, 2020; Martinet et al., 2021) use invariant causal prediction (ICP)

[1]Computer Vision Group, Freidrich Schiller University, Jena, Germany [2]German Aerospace Center (DLR), Institute for Data Science, Jena, Germany. Correspondence to: Wasim Ahmad <wasim.ahmad@uni-jena.de>.

Published at the ICML 2022 Workshop on Spurious Correlations, Invariance, and Stability. Baltimore, Maryland, USA. Copyright 2022 by the author(s).

for causal inference assuming that data samples are drawn from heterogeneous environments and can identify the ancestors of the response. These methods are limited in a sense that they estimate causal relationship of the candidate variable with the target only rather than discovering full causal graph as done in our work through multivariate target modelling. Moreover, they search for subset of predictors which remains unchanged throughout environments while we estimate a consistent causal set using model invariance across interventional environments over multiple forecast windows. ICP outputs the intersection of all subsets of invariant predictor which forms a strict subset of all direct causes or may even be empty. This is because disjoint sets of predictors can be invariant, yielding an empty intersection. In (Mogensen et al., 2022), the authors introduce and characterize minimally invariant sets of predictors, that is, invariant sets S for which no proper subset is invariant. They propose to consider the union S_{IAS} of all minimally invariant sets, where IAS stands for invariant ancestry search.

The method in (Peters et al., 2016) works with non-stationary data assuming that the environments are known in order to exploit the invariance. It is argued that without knowledge of the environments, the causal inference becomes difficult. Moreover, estimating the environments from data in naive manner e.g., clustering, change point detection, etc., and subsequently applying the existing methodology may lead to a loss of the methods coverage, guarantee or yield less powerful results. The authors of (Pfister et al., 2019) do not estimate the environments but instead utilizes the existing non-invariances in observations. In our method, we generate knockoff representation of the predictors to represent different environments and test model invariance for causal inference in nonlinear systems. Previously we used deep networks with Knockoffs-based interventions to estimate Granger causality in nonlinear multivariate time series (Ahmad et al., 2021).

The authors of (Tank et al., 2021; Romano et al., 2020) utilize deep networks combined with sparsity-inducing penalties on the weights to estimate causal significance of the predictors. However, it is difficult to retrieve clear information from weights of the deep networks that can be used for interpretability (Luo et al., 2020). The widely used PCMCI (Runge et al., 2019) method applies linear and nonlinear conditional independent (CI) tests to infer causality in multivariate time series. The extended version of cause-effect variational autoencoders (CEVAE) (Louizos et al., 2017) integrates the domain-specific knowledge for nonlinear causal inference in ecological time series (Trifunov et al., 2019). Time-frequency analysis has been done to handle spurious associations in time series (Shadaydeh et al., 2019) however, they use spectral representation of linear VAR-GC and might struggle in case of nonlinearity.

3. Methodological Background

In this section we introduce the necessary background on DeepAR and the Knockoffs framework.

3.1. DeepAR

DeepAR is a powerful method suitable for probabilistic forecasting of high dimensional nonlinear non-stationary time series (Salinas et al., 2020). It uses LSTM-based autoregressive RNNs (Graves, 2013) and learns nonlinear interactions from the history of all the related time-series. Few key attributes of DeepAR are: It extracts hidden features in the form of trends and seasonality in data, and characterizes conditioning on the entire set of variables during inference which can be helpful in identifying spurious relations. Moreover, lesser hand-crafted feature engineering is needed in DeepAR to capture complex, group-dependent behavior.

3.2. Knockoffs

The Knockoff framework is developed to estimate feature importance using CI testing with controllable false discovery rate (Barber & Candès, 2015). Given a set of observed variables $Z = (Z_1, \ldots, Z_n)$ in an environment with known distribution P_Z, and a predictive model, the knockoff copies of the observed variables, defined as $\widetilde{Z} = (\widetilde{Z}_1, \ldots, \widetilde{Z}_n)$, represent interventional environments while maintaining the in-distribution and decorrelation property of the observational environment. Knockoff samples are statistically null-variables, generated without knowing model response. Moreover, knockoffs should satisfy the pairwise exchangeability condition (Barber & Candès, 2015)

$$(Z_1, \widetilde{Z}_2, \widetilde{Z}_1, Z_2) \stackrel{d}{=} (Z_1, Z_2, \widetilde{Z}_1, \widetilde{Z}_2) \tag{1}$$

for any subset $\mathbb{A} \subseteq 1, \ldots, n$, here $\stackrel{d}{=}$ represents equal distributions. The $(Z, \widetilde{Z})_{swap(\mathbb{A})}$ is obtained from $(Z, \widetilde{Z})$ by swapping the entries Z_j and $\widetilde{Z}_j$ for each $j \in \mathbb{A}$.

The knockoff mechanism can be thought of as generating a probability distribution $P_{\widetilde{Z}|Z}(.|z)$ which is the conditional distribution of $\widetilde{Z}$ given $Z = z$ chosen such that the obtained joint distribution of $(Z, \widetilde{Z})$ being equal to

$$P_Z(z) P_{\widetilde{Z}|Z}(\widetilde{z}|z),$$

is pairwise $(z_j, \widetilde{z}_j)$ symmetric and satisfy the exchangeability condition in (1). For variables with Gaussian distribution $P_Z = \mathcal{N}_n(\mathbf{0}_n, \Sigma)$, (Candes et al., 2017) shows that knockoffs $\widetilde{\mathbf{Z}}_{i,*}$ can be drawn from the conditional distribution

$$P_{\widetilde{Z}|Z}(.|\mathbf{Z}_{i,*}) = \mathcal{N}_n((\mathbf{I}_n - S\Sigma^{-1})\mathbf{Z}_{i,*}, 2S - S\Sigma^{-1}S) \tag{2}$$

for any fixed diagonal matrix S satisfying $0 \leq S \leq 2\Sigma$.

This leads to joint distribution of $(\mathbf{Z}_{i,*}, \widetilde{\mathbf{Z}}_{i,*})$ being equal to

$$\mathcal{N}_{2n}\left(\mathbf{0}_{2n}, \begin{bmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{bmatrix}\right),$$

which satisfies the pairwise exchangeablity (1). Here S is to be maximized which represents the difference between the original variables and their knockoff version. As a result the off-diagonal term $\Sigma - S$ is minimized making the covariance of the original variable with its corresponding knockoff version as low as possible.

4. Causality using Model Invariance through Knockoffs

Let $z_i, i = 1, \ldots, N$ be the $N-$variate time series. Each time series $z_{i,t}, t = 1, \ldots, r$ is a realization of length r real-valued discrete stochastic process $Z_i, i = 1, \ldots, N$. Our goal is to estimate the causal relations among all variables and retrieve a full causal graph. Throughout our work, we assume stationarity and causal sufficiency which implies $Z_i, i = 1, \ldots, N$ is a stationary stochastic process, and the set of observed variables $Z_i, i = 1, \ldots, N$ includes all of the common causes of pairs in Z. An extension of the current work to deal with non-stationary settings is in progress.

We exploit the idea of model invariance for causal discovery in nonlinear time series where we determine that the underlying causal structure of a system remains invariant over interventional environments. We model nonlinear time series with DeepAR and generate knockoff copies of the original data to test model invariance. A model is regarded invariant if the distribution, i.e., μ and σ^2, of the response residuals given its causal predictors do not change against interventional environments. The predictors are considered as non-causal if the residual distribution of the model response does not significantly change by exposing it to their knockoff representations. We retrieve causal graph with a consistent set of variables using hypothesis testing over residual distribution from multiple forecast windows.

To generate knockoff variables, we implement DeepKnockoffs (Romano et al., 2020), a framework for sampling approximate knockoffs. We use mixture of Gaussian models as implemented in (Gimenez et al., 2019) to obtain knockoff samples. The procedure consists of first sampling the mixture assignment variable from the posterior distribution. The knockoff variables are then sampled from the conditional distribution of the Knockoffs given the original variables and the sampled mixture assignment such that the necessary conditions for Knockoffs are satisfied. Swapping any knockoff variable with their original version must leave the underlying joint data distribution invariant as in (1). Given observed variable $Z \sim \mathcal{N}(\mu, \Sigma)$, we aim to build its knockoff copy $\widetilde{Z}$ in such a way that the joint distribution of $(Z, \widetilde{Z}) \sim \mathcal{N}(*, **)$ is Gaussian and

$$* = \begin{bmatrix} \mu \\ \mu \end{bmatrix}, ** = \begin{bmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{bmatrix},$$

where S is a diagonal matrix such that $** \succeq 0$. S represents the discrepancy between originals and knockoffs. We have the same μ for both original and knockoff version. The Knockoff framework maintain the same covariance among variables in set of generated Knockoffs as in originals which is represented by the diagonal entries in given covariance matrix. However, the covariance $\Sigma - S$ of the original variable with its corresponding knockoff copy in the off-diagonals is kept as low as possible by maximizing S. We obtain Knockoffs $\widetilde{Z}_j, j \subseteq 1, \cdots, n$, for all variables in the multivariate time series and substitute one variable Z_j at a time with $\widetilde{Z}_j$ for each $j \subseteq 1, \cdots, n$ to test model invariance and estimate cause-effect relationships.

To estimate causal effect of $z_i, i = 1, \ldots, N$ on the response z_j, we test model invariance using its forecast residuals by feeding a knockoff representation $\widetilde{z}_i$ of the original variable z_i. We obtain residuals $e_1, e_2, ..., e_n \sim R$ and $\tilde{e}_1, \tilde{e}_1, ..., \tilde{e}_n \sim \tilde{R}$ over forecast windows w in range $[20 - 30]$ with a step size of $[5 - 10]$ for each realization $z_{j,t}$ of the stochastic process $Z_j, j = 1, \ldots, N$ with and without intervention on $z_i, i \neq j$. R represents distribution of model residuals for actual outcome and $\tilde{R}$ shows residual distribution after intervention. We compare these two distributions using Kolmogorov–Smirnov (KS) test (Smirnov, 1939) for causal inference. The test statistic uses the supremum distance $D_{n,m}$ between R and $\tilde{R}$. Here n, m are the number of samples in the these two distributions.

$$D_{n,m} = \sqrt{\frac{nm}{n+m}} sup|R_n - \tilde{R}_m| \tag{3}$$

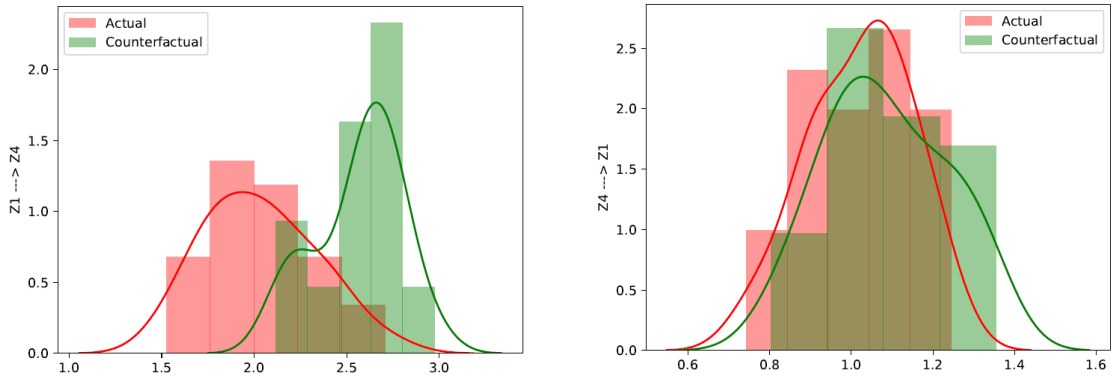
The null hypothesis H_0: z_i does not cause z_j, is accepted if the distribution of the residuals is approximately identical across environments $\tilde{R} \overset{d}{\approx} R_i$ with a significant level α in range $[5 - 10]\%$. In case the residual distribution is significantly different $\tilde{R} \overset{d}{\neq} R_i$, the alternate hypothesis H_1: z_i causes z_j is accepted. $D_{n,m}$ tends to have significantly larger value in alternate case.

5. Experimental data and Results

We conducted experiments on real and synthetically generated time series data. The synthetic data model has the generic form:

$$Z_t^j = a_j Z_{t-1}^j + \Sigma_i c_i f_i(Z_{t-\tau_i}^i) + \eta_t^j \tag{4}$$

for $j \in 1, ..., N$. The correlation coefficients a_j are selected from range $[0.2 - 1.0]$. Besides autodependency in



(a) Causal case (b) Non-causal case

Figure 1. (a) Model residual distribution for causal scenario where Z_1 causes Z_4 in our generated causal graph. (b) Model residual distribution for non-causal scenario where Z_4 is not causal for Z_1 based on ground truth topology.

the model, there exist linear and exponential functional dependencies $f_i(z)$. We included noise η_t^j, which represents data from uncorrelated, normally distributed noise with zero mean and variances in the range $[0.30 - 0.9]$. The coupling coefficients c_i and lags τ_i between variables are set to $[0.2 - 1.0]$ and $[0 - 10]$ respectively. To test our method, we generate multiple causal graphs with $N = 5$ nodes by changing functional dependencies, causal edges, time delays, dynamic noise and correlation coefficients throughout our experiments. To quantify the performance of applied methods, we calculate the following false positive rate (FPR) and F-score metrics in a variety of generated causal graphs.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{5}$$

$$\text{F-score} = \frac{\text{TP}}{\text{TP} + 0.5(\text{FP} + \text{FN})} \tag{6}$$

Here, TP is the number of causal links that exist in the ground truth causal graph and also detected by the applied methods; TN is the number of causal links that neither exist in the causal topology nor detected by the causality techniques; FP is the number of causal links that do not exist in the causal graph but detected by the methods and FN is the number of the causal links that exist in the causal graph but not detected by the used methods.

To show the effectiveness of proposed approach where we utilize Knockoffs-based intervention variables, we compare its performance against other widely used causal inference methods. Besides that, to highlight the power of using Knockoffs as intervention variables, we test the proposed approach with the following methods for generating intervention variables.

Distribution mean: The predictor z_i is replaced with $\overline{z}_i$, that is comprised of the mean of the original variable and Gaussian noise. $\overline{z}_i = \frac{1}{r}\sum_{t=1}^{r} z_{i,t} + \mathcal{N}(0, \Sigma)$.

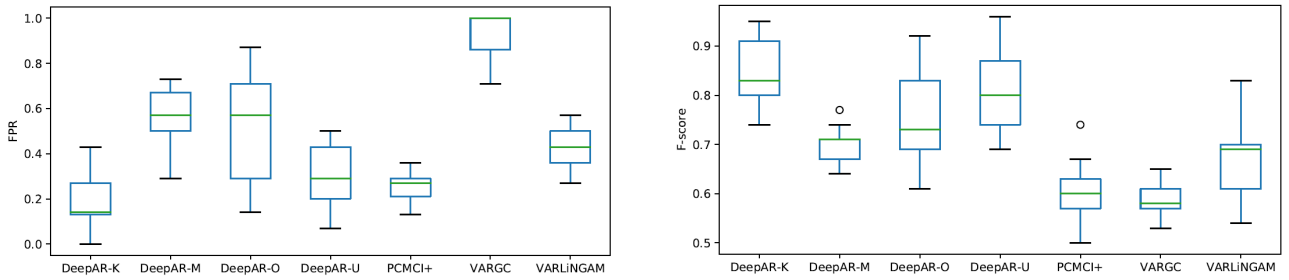
Uniform distribution: The original variable z_i is replaced by a variable z_i' having a uniform distribution ranging from minimum to maximum values of the original variables z_i as suggested in (Jiwoong Im et al., 2021).

Out-of-distribution: The variable z_i having a distribution D_i is replaced with another variable sampled from OOD $\overline{D}$ such that $\overline{D} \overset{d}{\neq} D_i$, different μ and σ^2. Moreover, it is generated to be uncorrelated with the original variables.

As a real-world application, we conducted experiments on average daily discharges of rivers in the upper Danube basin, provided by the Bavarian Environmental Agency [1]. We use measurements from 2017 to 2019 from the Iller at Kempten K_t, the Danube at Dillingen D_t, and the Isar at Lenggries L_t. The Iller flows into the Danube within a day, which implies an instantaneous causal link $K_t \rightarrow D_t$ and no direct causal links between the pairs K_t, L_t and D_t, L_t. Moreover all variables may be confounded by rainfall or other weather conditions (Gerhardus & Runge, 2020), which test the ability of the applied methods to distinguish between spurious associations and true causal relationships.

Results: We generate multiple causal graphs using synthetic data model and demonstrate the results in terms of FPR and F-score for VAR-GC, VARLiNGAM, PCMCI$^+$ and DeepAR with all four types of intervention variables represented as DeepAR-K for Knockoffs, DeepAR-O for OOD, DeepAR-M for mean and DeepAR-U for uniform distribution. We infer causality among nodes of the graph using model invariance through knockoff interventions where we use the KS test (3) to compare the residual distributions of

[1] https://www.gkd.bayern.de



(a) False Positive Rate

(b) F-score

Figure 2. (a) False positive rate (FPR) of used methods for causal inference in synthetic multivariate nonlinear time series, lower is better. (b) F-score of all applied methods for estimating causality in synthetic multivariate nonlinear time series, higher is better.

Table 1. Causality in river discharges time series data by VAR-GC, VARLiNGAM, PCMCI$^+$ and DeepAR with applied interventions types. The detection of a causal link between variables is represented by ✓ and absence of causality is shown by ×.

Pairs	Expected links	VAR-GC	VAR-LiNGAM	PCMCI$^+$	DeepAR-Knockoffs	DeepAR-OOD	DeepAR-Mean	DeepAR-Uniform
$K_t \to D_t$	✓	✓	✓	✓	✓	✓	×	✓
$K_t \to L_t$	×	×	✓	×	×	×	×	×
$D_t \to K_t$	×	×	×	×	×	✓	×	×
$D_t \to L_t$	×	✓	×	×	×	×	✓	✓
$L_t \to K_t$	×	✓	✓	✓	×	×	✓	×
$L_t \to D_t$	×	✓	×	×	×	×	✓	×

the model with and without interventions. An example is shown in Fig 1, where we illustrates the causal decision by our method between pair of nodes Z_1 and Z_4. Based on the ground truth from our generated causal graph, Z_1 causes Z_4 while the reverse in not true. Our method accurately identifies $Z_1 \to Z_4$, a true positive and $Z_4 \nrightarrow Z_1$, a true negative. Overall DeepAR-K performs better than others in accurate causal discovery. Since the Knockoffs-framework is originally developed to control false discoveries, we can see low FPR for Knockoffs-based interventions as shown in Fig 2 (a) compared to other methods and intervention types. At the same time, the power of DeepAR-K to identify true causal links is high as illustrated in Fig 2 (b), which could be attributed to the uncorrelated property of the Knockoffs with their originals. The closest to our proposed DeepAR-K in terms of achieving high F-score and low FPR is DeepAR-U. PCMCI$^+$ also achieves relatively better FPR but shows low power for true causal detection along with VAR-GC and VARLiNGAM. The low detection power of these methods might be because of nonlinearity in the generated causal graphs.

For river discharges data, we have one contemporaneous link $K_t \to D_t$ as per description in the dataset, and we expect the applied methods to identify it. Our proposed method along with all other methods except DeepAR-Mean accurately detected $K_t \to D_t$ as shown in Table 1. Moreover, our method does not report any false causal link. However, the rest of the methods show spurious associations between various pairs of river discharges time series. DeepAR-OOD incorrectly detected the reversed causal link $D_t \to K_t$ too. VAR-GC and DeepAR-M identified a bidirectional link between D_t and L_t which may be because of the weather acting as a confounder since there is no expected direct causal relationship. DeepAR-U, DeepAR-OOD and PCMCI$^+$ have one false link detection each in different pairs.

6. Conclusion

We proposed to test model invariance across interventional environments over multiple forecast windows to discover full causal graph in multivariate nonlinear systems. DeepAR is used to learn nonlinear relationship in multivariate time series. The model invariance is tested via Knockoffs-based interventions for causal inference. We used synthetic and real time series data to evaluate applied methods. The results indicate that the estimation of the environments using Knockoff framework to test model invariance for causal inference yield better results in comparison to other intervention types. Results also show that the proposed method outperforms VAR-GC, VARLiNGAM and PCMCI$^+$, however having a higher computation time. Further we aim to extend the current work for non-stationary settings and test our method on other real-world time series.

Acknowledgements

This work is funded by the Carl Zeiss Foundation within the scope of the program line "Breakthroughs: Exploring Intelligent Systems" for "Digitization — explore the basics, use applications" and the DFG grant SH 1682/1-1.

References

Ahmad, W., Shadaydeh, M., and Denzler, J. Causal inference in non-linear time-series using deep networks and knockoff counterfactuals. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 449–454. IEEE, 2021.

Barber, R. F. and Candès, E. J. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5), Oct 2015. ISSN 0090-5364. doi: 10.1214/15-aos1337. URL http://dx.doi.org/10.1214/15-AOS1337.

Candes, E., Fan, Y., Janson, L., and Lv, J. Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection, 2017.

Gamella, J. L. and Heinze-Deml, C. Active invariant causal prediction: Experiment selection through stability. *Advances in Neural Information Processing Systems*, 33: 15464–15475, 2020.

Gerhardus, A. and Runge, J. High-recall causal discovery for autocorrelated time series with latent confounders. *34th Conference on Neural Information Processing Systems*, 2020.

Gimenez, J. R., Ghorbani, A., and Zou, J. Knockoffs for the mass: new feature importance statistics with false discovery guarantees. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2125–2133. PMLR, 2019.

Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

Graves, A. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

Heinze-Deml, C., Peters, J., and Meinshausen, N. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2), 2018.

Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11(5), 2010.

Jiwoong Im, D., Cho, K., and Razavian, N. Causal effect variational autoencoder with uniform treatment. *arXiv e-prints*, pp. arXiv–2111, 2021.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Louizos, C., Shalit, U., Mooij, J., Sontag, D., Zemel, R., and Welling, M. Causal effect inference with deep latent-variable models. *arXiv preprint arXiv:1705.08821*, 2017.

Luo, Y., Peng, J., and Ma, J. When causal inference meets deep learning. *Nature Machine Intelligence*, 2(8):426–427, 2020.

Martinet, G., Strzalkowski, A., and Engelhardt, B. E. Variance minimization in the wasserstein space for invariant causal prediction. *arXiv preprint arXiv:2110.07064*, 2021.

Mogensen, P. B., Thams, N., and Peters, J. Invariant ancestry search. *arXiv preprint arXiv:2202.00913*, 2022.

Peters, J., Bühlmann, P., and Meinshausen, N. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):947–1012, 2016.

Pfister, N., Bühlmann, P., and Peters, J. Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, 114(527):1264–1276, 2019.

Romano, Y., Sesia, M., and Candès, E. Deep knockoffs. *Journal of the American Statistical Association*, 115(532): 1861–1872, 2020.

Runge, J. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1388–1397. PMLR, 2020.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11):eaau4996, 2019.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

Shadaydeh, M., Denzler, J., García, Y. G., and Mahecha, M. Time-frequency causal inference uncovers anomalous events in environmental systems. In *German Conference on Pattern Recognition*, pp. 499–512. Springer, 2019.

Smirnov, N. V. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2):3–14, 1939.

Tank, A., Covert, I., Foti, N., Shojaie, A., and Fox, E. B. Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021. ISSN 1939-3539. doi: 10.1109/tpami.2021.3065601. URL http://dx.doi.org/10.1109/TPAMI.2021.3065601.

Trifunov, V. T., Shadaydeh, M., Runge, J., Eyring, V., Reichstein, M., and Denzler, J. Nonlinear causal link estimation under hidden confounding with an application to time series anomaly detection. In *German Conference on Pattern Recognition*, pp. 261–273. Springer, 2019.