Cardiff Economics Working Papers

Working Paper No. E2023/10

Chameleon models in economics: A note

Michael Hatcher and Patrick Minford

April 2023

ISSN 1749-6010

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
business.cardiff.ac.uk

Chameleon models in economics: A note

Michael Hatcher, University of Southampton, Patrick Minford, Cardiff University

April 2023

Abstract
In a recent article Pfleiderer (2020) argues that models should be put through a 'reality filter' before they should be taken seriously as a basis for empirical testing or policy analysis. He regards models that make dubious assumptions as 'chameleons', because while as theory they may give insights, their empirical performance is artificial and disguises their true nature as impossible beasts. In this note we argue that the correct response to such models is not to reject them *ex ante* based on a 'reality filter' but to subject them to rigorous empirical testing wherever possible. We describe a methodology in macroeconomics that can be used for this purpose, and we show how this approach avoids several problems that arise from using a 'reality filter' without further scrutiny.

1. Introduction

Theoretical models in economics and finance are based on simplifying assumptions. While such assumptions are necessarily "unrealistic" in a descriptive sense, conventional wisdom holds that some assumptions are more plausible than others and it is well-known that different assumptions can have substantive implications for outcomes or policy conclusions derived from economic models. Following this lead, several recent papers have questioned whether standard assumptions in economic models are fit for purpose.[1] Most notably, Pfleiderer (2020) argues that some theoretical models in economics and finance are 'chameleons' because they are built on assumptions with dubious connections to the real world yet are used as a basis for understanding our economy and drawing policy conclusions. According to this view, chameleons are nurtured by the reluctance of researchers to judge models by their assumptions and the difficulty of conducting empirical tests. However, Pfleiderer also argues the solution is simple: we should run our models through a 'reality filter' and reject models whose assumptions are strongly at odds with our priors.

In this note, we argue against using a 'reality filter' to reject certain models or theories. Our main criticisms are that such an approach is inconclusive for models of interest and neglects empirical tests that could be useful in identifying 'chameleon' models. We propose an alternative solution to the 'chameleons' problem based on rigorous empirical testing of models; our argument here is that such testing should show chameleon models for what they are, while retaining those models that might on the surface look like chameleons but are in fact useful for policy analysis (potentially after some modification). We describe an empirical methodology in macroeconomics that can be used for this purpose, and we explain how this approach would avoid several problems with Pfleiderer's recommendation of using a 'reality filter' without further scrutiny.

Our paper is built around two main arguments. First, researchers do routinely use filters to judge economic models, but these are useful primarily in *clear-cut* cases such as 'nonsense' models. By contrast, 'chameleons' are models that make `dubious' – though not absurd – assumptions and are taken seriously (by some) as a basis for understanding the economy and drawing policy conclusions. These features make chameleons hard to 'catch' with our filters because, unlike 'nonsense models' which are atheoretical or logically incoherent, their flaws are not obvious. Since our filters are inconclusive in such cases, they will be of little use in the absence of further evidence. Second, we argue that attempts to 'filter out' chameleon models, as suggested by Pfeiderer, are likely to be

[1] See, for example, Colander et al. (2009), Gilboa et al. (2014), Ng (2016), and Vines and Willis (2018).

inferior to empirical testing when the latter is possible. Indeed, if we see a potential 'chameleon' model which makes assumptions that appear to have dubious connections to the real-world, then rather than assuming we possess the knowledge and judgement to correctly classify the model, shouldn't we subject the model to tests that could (in principle) reveal the model's true nature? In short, we argue that for non-trivial models, we will need empirical testing to make progress.

We describe a methodology in macroeconomics that can be used for this purpose, and we show how such empirical testing could avoid several problems with using a 'reality filter' to reject models or theories without any further scrutiny. Our suggested approach is based on work in macroeconomics that formally tests models in the Popperian sense that they will be rejected if they are unable to mimic the dynamic behaviour of the model variables in the data. Importantly, this approach offers multiple ways to test models against data, such that the scope for subjective bias and *misuse* of models should be reduced. Further, it is possible to test sub-parts of a model and identify some sources of misspecification, which may yield useful information about where falsified models require modification and improvement. Thus, while we agree with Pfleiderer that 'chameleons' can be nurtured by bad practice, our proposed solution is very different to his informal 'reality filter'.

2. Chameleons revisited

Pfleiderer (2020) defines a 'chameleon' as an economic model which "is built on assumptions with dubious connections to the real world but nevertheless has conclusions that are uncritically (or not critically enough) applied to understanding our economy" (p. 81). He invites us to think of chameleons in relation to 'bookshelf models' that may provide insights – such as showing the implications of a particular set of assumptions – but are not intended to be applied to the real world. In such models, the assumptions are often 'dubious' in the sense that the assumed problem may have no clear real-world counterpart.[2] A chameleon is then defined as a 'bookshelf model' whose conclusions are applied uncritically for understanding the real world (as though it is a 'policy model'); yet when the model and its assumptions are criticised, it is defended as a bookshelf model and hence avoids the 'reality filter'. In other words, 'chameleons' avoid the real-world filter because when their assumptions are criticised as 'dubious' it is said that no defence is required.

The question is thus: how to avoid 'chameleons' while retaining models that are potentially useful? Clearly, if we could recognise 'chameleons' simply by passing their assumptions through a 'reality filter', then we could separate out these 'bookshelf' models from others and avoid using them for policy analysis. But can we really do so? Pfleiderer suggests that we can recognise 'chameleons' by comparing our (non-empty) set of prior beliefs to the assumptions made in theoretical models.

We would dispute this thesis vigorously. The key difficulty with such a filter is that the problem of modelling is that no one knows what the true model – i.e. 'reality' – is. Since 'chameleons' are models with dubious (but not absurd) assumptions, it is not obvious which models should pass through the filter and which should not. These models can be contrasted with 'nonsense' models which clearly fail to pass through the relevant filters and which, trivially, avoid misclassification and *misuse* because they are *not* used. Thus Pfleiderer gives the example of an inflation theory based on the height of Oscar winners, which he says would be discarded by a reality filter. But surely it would never be proposed as a model as it is not based on economic theory; it could be said to be discarded by a 'theory filter', in the sense that no sensible economic theory gives rise to it. The same goes for

[2] For instance, Pfleiderer (2020, p. 83) gives the example of an optimal contract in a principal-agent problem that is not available in practice and could not be made available. Pfleiderer also gives two examples of models from finance which he considers to be 'chameleons' and hence bookshelf models; see Section 3 in the paper.

other 'nonsense' models, which are atheoretical or logically incoherent, or for 'toy models' intended as bookshelf models. In all these cases, a reality filter would make little difference because we *already* exclude such models from the set of models suitable for *real-world* analysis.

For models without such obvious flaws, no easy classification is possible. Recall that Pfleiderer argues that 'chameleons' are defended as valid 'bookshelf' models – and hence they are not logically incoherent or atheoretical, though they may be 'toy' models. What is it, then, that separates a bookshelf model from one that is suitable for policy analysis? The answer, according to Pfleiderer, is that the latter models have a tighter connection with the real world, though exactly what this means remains unclear. In short, the difficulty for researchers with potential 'chameleon models' to hand is to decide, in each case, whether the reality filter has been passed. Unfortunately, when we are looking outside of 'nonsense' models (and in empirical modelling we always are) this decision is non-trivial, and hence a reality filter is likely to be of little use.

Given these problems with a 'reality filter', how should we go about detecting 'chameleons'? We argue that potential 'chameleons' should be subjected to rigorous empirical testing. The basic idea is quite simple: since chameleons arise due to *misuse* of models (rather than weaknesses of modelling *per se*), it makes sense to formalise the process of detecting such models, thus reducing the scope for subjective factors to play a role. We do not claim, of course, that all empirical testing is definitive or that such testing is infallible (especially since decisions about *what* to test must be made by researchers). However, an important advantage is that these methods can be used to formally test whether a model can mimic the dynamic behaviour of the data it was intended to explain.[3]

The empirical methodology we put forward offers several ways to formally test models against the data in a Popperian sense. Further, these methods can be used to test sub-parts of a wider model, which may bring to light areas in which falsified models can be modified to improve their empirical performance. We argue these methods should be used to test our models and theories, with the aim of revealing some models as 'chameleons' (i.e. bookshelf models), while retaining others that may superficially look like 'chameleons' but can be useful for drawing real-world conclusions or policy analysis (potentially after some modifications). We now describe this empirical testing approach.

3. A formal alternative

Models propose hypotheses about how agents behave and what constraints they face. As noted, Pfleiderer does not define what the reality is that would 'filter' hypotheses; or who would determine what is 'realistic', implying that it is obvious and can be imposed by anyone as a prior belief. Yet if it is undefined, it cannot either be obvious or a valid methodology. We argue that hypotheses should be tested by a powerful *empirical* method to see if they can mimic the data behaviour. All we can do in this short paper is set out a viable empirical methodology for testing economic models; and draw attention to the dangers of departing from it in the ways suggested by Pfleiderer. It is in a good empirical testing methodology that we see an appropriate 'filter' that theories must pass.[4]

We begin with model specification. Here a candidate hypothesis is developed from economic theory. Typically, economic agents maximise utility subject to their constraints, including assumptions about

[3] Plainly, models should be tested only on the data they are meant to explain. For example, models may only apply to certain episodes because they embody policy regimes that only applied to those episodes or practices (such as high price or wage rigidity) that were particularly prevalent at that time and place. As noted below, this feature does not make sound inference infeasible for empirical methods that work on small samples.

[4] Although we focus our comments below on a particular testing methodology (namely indirect inference), it should be understood that our conclusions will apply to other testing methods that have similar properties (e.g. high power in small samples and robustness to some forms of model misspecification).

their information, expectations, and utility functions. It is this stage of model building that 'weeds out' nonsense models, such as Pfleiderer's inflation theory based on the height of Oscar winners. We are thinking here of dynamic macroeconomic models in the spirit of Smets and Wouters (2003, 2007); however, the models we have in mind are *not* restricted to rational expectations or linearity as in their papers. The resulting model must be identified: i.e. have a unique reduced form.

Pfleiderer seems to have a low opinion of the ability of empirical testing to reject theories that cannot mimic reality. This view appears to underlie his argument that we should apply a reality filter, effectively as a substitute for empirical testing of theories. Yet we know that powerful methods exist, even in small samples. Le et al (2016) and Meenagh et al (2019) survey these methods and show that indirect inference (Smith,1993; Gourieroux et al, 1993, 1996) supplies a low-bias estimator and a highly powerful test in small samples; in large samples this approach is equivalent to full-information maximum likelihood (FIML). By 'small samples' we refer not only to the absence of 'long' time series (for example, GDP was not measured until the 1930s), but also to potential shifts in the data-generating processes under different policy regimes, such that model estimation and evaluation might usefully be restricted to certain 'regimes' for which a particular model is relevant.[5]

To test whether a hypothesised model can be rejected as the true model, the indirect inference method proceeds as follows. We first describe the data behaviour by an *auxiliary model* which is typically a reduced-form vector autoregression (VAR), but can also take other forms, such as moments or impulse response functions.[6] The model is then simulated to generate numerous samples on each of which the auxiliary model is estimated, so creating a joint distribution of its parameters according to the model being tested. From this the probability of the data-based parameters can be derived, providing a Wald test of the model. The model's structural parameters are estimated as those giving the lowest Wald value, and the estimated model is then tested on this value. The Wald test is then a test of the null hypothesis that the model describes the data; low values of the Wald statistic favour non-rejection of the model (i.e. the null hypothesis).

Monte Carlo experiments in Le et al (2016) and Meenagh et al (2019) show that in small samples this method has low estimation bias and high power as a test against numerically inaccurate models. Notably, the indirect inference approach is found to have higher power than FIML in small samples (likelihood ratio tests), though the latter approach is more widely used in the literature. Recent work by Meenagh et al (2022) has also shown that Bayesian priors create estimation bias if inaccurate; and will also be rejected in this case. We therefore recommend that models should be tested inclusive of their priors to guard against this problem. Only if priors are known to be true with certainty should they be imposed in estimation and testing; thus, Pfleiderer's suggested approach of imposing strong priors is supported only if we know a lot about the 'true model'.

While Pfleiderer uses the term 'priors' as a general term for the assumptions of a model, in macroeconomic models the term is usually taken to mean Bayesian priors on the model parameters. In estimation researchers typically specify prior distributions for the estimated structural parameters of the model, with 'uninformative priors' corresponding to (e.g.) a non-degenerate uniform distribution, whereas a degenerate prior distribution implies that a specific prior belief is held with certainty (i.e. taken as exactly true). However, it is important to note that *different sets of assumptions* can be nested via appropriately defined model parameters; for example, one parameter could be the relative weight on rational expectations versus adaptive learning, with a

---

[5] See, for example, the paper by Benati (2008) on different inflation regimes and our comment in Fn. 3.

[6] Fernández-Villaverde et al. (2007) give conditions such that (in population) one may infer economic shocks and impulse responses of a theoretical model from the innovations and impulse responses of a VAR.

uniform prior on the interval [0,1], such that estimation may be informative about which expectational hypothesis performs better.

As examples of this approach, Le et al. (2011) test a version of the Smets and Wouters (2007) model with a relative weight on flexible versus 'sticky' wages and prices, while Levine et al. (2012) estimate a model with imperfect information about current state variables and compare model fit against a full-information rational expectations version which allows some fraction of agents to have adaptive expectations. In a similar vein, Milani (2007) and Slobodyan and Wouters (2012) study how adaptive learning performs vis-à-vis a rational expectations benchmark. From these examples we see that the empirical methodology described above may be used to test a wide range of modelling assumptions.

This last point is highly relevant to the selection of assumptions, which are frequently imposed as priors. For instance, some modellers set out micro-founded assumptions based on what they take to be observed micro reality, then impose them as priors and so fail to test them on the data. Rather than checking them via a 'reality filter' we argue that they should be tested empirically as proposed – i.e. we should acknowledge our ignorance by imposing 'weak' priors. Pfleiderer's suggested approach is at the opposite end of spectrum: it assumes that researchers possess non-trivial knowledge of the true model. If they do not, then the consequences for inference can be dire.

We close by noting an extra feature of these methods which is attractive from a model evaluation perspective. In model evaluation we will sometimes find that rejection of a model is marginal; in such cases, it would be useful to have some idea about which part(s) of the model might require modification to improve the model. Recent work facilitates this process by allowing researchers to test sub-parts of a wider model for misspecification; for example, Inoue et al. (2020) provide a general method for detecting and locating sources of model misspecification in theoretical models,[7] whereas Minford et al. (2019) show how one may test a sub-part of a larger model using the method of indirect inference. With these methods in hand, modellers are not confined to 'shooting in the dark' but may get useful feedback on potential weaknesses in existing models and theories.

4. Conclusion

In summary, we have argued here that the correct response to chameleon models is not a 'reality filter' based on prior beliefs, suggested by Pfleiderer (2020). Such a filter will catch nonsense or 'toy' models, but for most models of interest the distinction between 'realistic' and 'dubious' assumptions is not clear-cut, rendering such a filter of little use. Some are bound to disagree with us about the problems with a reality filter, but this need not result in stalemate. We argue that the best response to suspected 'chameleons' is to subject them to empirical tests which could reveal their true colours as 'bookshelf' models. In this way economic modelling may benefit from criticisms which are targeted at dubious assumptions, provided that these critiques are followed up by empirical tests.

References

Benati, L. (2008). Investigating inflation persistence across monetary regimes. *The Quarterly Journal of Economics*, *123*(3), 1005–1060.

Colander, D., Goldberg, M., Haas, A., Juselius, K., Kirman, A., Lux, T., & Sloth, B. (2009). The financial crisis and the systemic failure of the economics profession. *Critical Review*, *21*(2-3), 249-267.

[7] See also the paper by Del Negro and Schorfheide (2009).

Del Negro, M. and Schorfheide, F. (2009). Monetary policy analysis with potentially misspecified models, *American Economic Review* 99, 1415-50.

Fernández-Villaverde, J., Rubio-Ramirez, F., Sargent, T. and Watson, M. (2007). ABCs (and Ds) of Understanding VARs, *American Economic Review*, Vol. 97, pp. 1021–1026.

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2014). Economic models as analogies. *The Economic Journal*, *124*(578), F513-F533.

Gourieroux, C. and Monfort, A. (1996). Simulation Based Econometric Methods, CORE Lectures Series, Oxford University Press, Oxford.

Gourieroux, C., Monfort, A. & Renault, E. (1993). Indirect inference, *Journal of Applied Econometrics*, Vol. 8(S), pp.85-118.

Inoue, A., Kuo, C. H., & Rossi, B. (2020). Identifying the sources of model misspecification. *Journal of Monetary Economics*, 110, 1-18.

Le, M., Meenagh, D., Minford, P. & Wickens, M. (2011). How much nominal rigidity is there in the US economy? Testing a new Keynesian DSGE model using indirect inference, *Journal of Economic Dynamics and Control*, Vol. 35(12), pp.2078-2104.

Le, M., Meenagh, D., Minford. P., Wickens, M. & Xu, Y. (2016). Testing Macro Models by Indirect Inference: A Survey for Users, *Open Economies Review*, Vol. 27(1), pp.1--38.

Levine, P., Pearlman, J., Perendia, G., & Yang, B. (2012). Endogenous persistence in an estimated DSGE model under imperfect information. *The Economic Journal*, *122*(565), 1287-1312.

Meenagh, D., Minford, P., Wickens, M., & Xu, Y. (2019). Testing DSGE models by indirect inference: a survey of recent findings. *Open Economies Review*, *30*, 593-620.

Meenagh, D., Minford, P. & Wickens, M. (2022). The Macroeconomic Controversy Over Price Rigidity — How to Resolve it and How Bayesian Estimation has Led us Astray, *Open Economies Review*, 2022, vol. 33, issue 4, No 1, 617-630.

Milani, F. (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics*, *54*(7), 2065-2082.

Minford, P., Wickens, M., & Xu, Y. (2019). Testing part of a DSGE model by indirect inference. *Oxford Bulletin of Economics and Statistics*, 81(1), 178-194.

Ng, Y. K. (2016). Are unrealistic assumptions/simplifications acceptable? Some methodological issues in economics. *Pacific Economic Review*, *21*(2), 180-201.

Pfleiderer, P. (2020). Chameleons: The misuse of theoretical models in finance and economics. *Economica*, *87*(345), 81-107.

Slobodyan, S. and Wouters, R. (2012). Learning in an estimated medium-scale DSGE model. *Journal of Economic Dynamics and Control*, *36*(1), 26-46.

Smith, A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions, *Journal of Applied Econometrics*, Vol. 8(S), pp.63-84.

Vines, D., & Wills, S. (2018). The rebuilding macroeconomic theory project: an analytical assessment. *Oxford Review of Economic Policy*, *34*(1-2), 1-42.