The Financial Econometrics of Price Discovery and Predictability

Seema Narayan and Russell Smyth*

Abstract:
This paper reviews recent econometric developments in the literature on price discovery and price predictability. For both areas, we discuss traditional approaches to econometric modeling, limitations to these approaches, and recent developments designed to overcome them. We also discuss the state-of-the-art and suggest future research. Three main conclusions are drawn. First, while many recent empirical applications in price discovery and price predictability are on the frontier of econometric methods, further developments are needed to increase relaxation of relevant assumptions and push the boundaries of applications. Second, future research in econometric modelling needs to combine/synthesize recent developments across multiple econometric issues rather than proceeding in a piecemeal manner, for instance, by integrating developments in the time series literature into panel-based frameworks. Third, recent econometric literature is generating findings that challenge long-held beliefs about apparent empirical regularities in price discovery and price predictability, thus presenting opportunities to develop relevant theory.

JEL Classification Numbers: C53, C58, G12, G13, G14

* We thank Paresh Narayan for helpful suggestions on an earlier version of this paper.

© 2015 Seema Narayan and Russell Smyth
All rights reserved. No part of this paper may be reproduced in any form, or stored in a retrieval system, without the prior written permission of the author.
1. Introduction
The primary motive for investing in domestic or international stocks is to maximize returns. This, however, encompasses risk, as both return and risk are conceptually tied to the price of the stock. Given this, an understanding of price formation is fundamental to achieving higher returns and lowering risks. It is therefore not surprising that scholars and practitioners in the field of financial economics are engaged in research that deals with the ‘what’ and ‘how’ of current and future price formation. The following survey examines the role of econometrics in this venture, and broadly covers financial econometric developments in the areas of price discovery and price predictability.

The main reason we focus on price discovery and price predictability together is that conceptually these two areas of financial economics are related. Dolatabadi et al. (2014) show that if one can ascertain that in a two-market (say X and Y) setup, market X dominates market Y in the price discovery process, the implication is that one can use market X to predict (and also forecast) market Y. To put it differently, the market with more information can be used to forecast the market with less information. Price discovery is a first step because it tells us that market X has more information content than market Y because the former dominates the price discovery process.

A further reason for considering price discovery and price predictability in the same review is that not only are these fields of financial economics conceptually related, but in many respects both areas are informed by similar time series econometric developments, particularly with respect to integration and cointegration. For example,
both areas have been informed by developments in unit root and cointegration testing designed to derive added power, such as the use of panel-based models. Similarly, both areas have been advanced by ongoing econometric developments designed to address non-linearities and structural change in data.

The catalyst for many recent econometric developments in both areas is the theoretical insights of Figuerola-Ferretti and Gonzala (2010). Building on the components share (CS) methods and a simultaneous price dynamics model proposed by Garbade and Silber (1983), Figuerola-Ferretti and Gonzala (2010) propose a theoretical model of price discovery characterized by finite elasticity of supply of arbitrage services and endogenous convenience yields. Their model demonstrates a perfect link between an extended Garbade and Silber (1983) framework and the CS approach, suggesting that the cointegrated vector autoregressive (CVAR) model (Johansen, 1995) is the natural empirical model to test for price discovery.

This insight is subsequently extended by Dolatabadi et al. (2014a, 2015) to develop a fractionally cointegrated vector autoregressive (FCVAR) model to examine price discovery, with the authors showing how CVAR and FCVAR links price discovery and price predictability. Thus, econometrics developments in both areas are starting to come together, and it is likely that this convergence will continue in the future. Consequently, given that there is a theoretical basis to using CVAR/FCVAR to test for price discovery and price predictability based on Figuerola-Ferretti and Gonzala (2010) and extensions of their theoretical model, it makes sense to focus on the evolution of unit root/cointegration testing in both literatures and how these are linked.
It should be noted that in chronicling the development in financial econometrics in the period 1980-2000, Bollerslev (2001) argues that time-varying volatility models (such as ARCH/GARCH and stochastic volatility formulations) and robust methods-of-moments-based estimations procedures (such as Generalized Methods of Moments) stand out as milestones. We do not review these important developments, however, as there are already several broad textbook treatments (see e.g. Campbell et al., 1997) as well as surveys on ARCH/GARCH and stochastic volatility/realized volatility (see Bollerslev et. al., 1992, 1994; Engle, 1995; Ghysels et al., 1996; Shephard, 1996; Barndorff-Nielsen and Shephard, 2014; Bollerslev, 2010; see also Bollerslev and Hodrick, 1995; Pagan, 1996; Bollerslev, 2001).

The structure of the survey is as follows. For both topics the main areas covered will be symmetric, namely: motivation and implications; overview of existing studies; state of the art econometric techniques; and suggestions for future research.

2. Price Discovery

2.1 Motivation and implications

Price discovery refers to the means through which information is transformed into prices. The price discovery literature centers on the speed at which related markets reach the equilibrium asset price, with the market in which new information is reflected most rapidly considered to be the one with the price discovery function.

There are multiple factors motivating studies of price discovery in financial markets. Improved understanding of price discovery not only assists investors to make better-informed decisions, but because it involves knowing how information translates into price across closely related markets, it facilitates risk management (Westerlund et al.,
2014). Improved understanding of price discovery also facilitates more efficient asset evaluation. Evidence from experiments suggests that trading the same asset across two markets simultaneously increases the speed at which efficient prices are reached (Forsythe et al., 1982, 1984). Given this, better understanding of the transmission mechanism of information across markets, and the rate at which information is reflected in prices, can provide evidence on the relative efficiency of markets.

Studying price discovery also provides insights into how markets incorporate new information about asset values, as well as the relative importance of fundamentals and speculation in this process. For example, the lead-lag relationship between spot and future markets has implications for evaluating the role of speculation in driving commodity prices. If price discovery occurs in the futures market, this suggests that speculation is driving commodity prices; however, if price discovery occurs in the spot market, the implication is that changes in market fundamentals are driving demand and supply (hence prices) for the commodity (Kaufmann and Ullman, 2009).

Price discovery also has implications for the preferred markets in which informed traders invest. If price discovery occurs in futures markets, this implies that informed traders might prefer to trade in futures markets rather than spot markets.

2.2 Types of price discovery in financial markets

There are at least three related sets of studies in the price discovery literature. One set of studies estimates the information share of each market for fragmented securities traded across multiple markets (see e.g. Eun and Subherwal, 2003; Harris et al., 1995, 2002; Hasbrouck, 1995). There has been a surge in the number of studies of this sort, which at least in part reflects growing market fragmentation (De Jong and Schotman, 2010; Westerlund et al., 2014). The principal finding from these studies is that price
discovery predominantly occurs in the home market, and that prices in the foreign market adjust to the home market. Notwithstanding, evidence exists that the relative importance of home and foreign market may change over time, particularly if there have been financial crises (Caporale et al., 2014; Fernandes and Scherrer, 2014).

A second set of studies estimates the information share of each market across interconnected markets such as spot and futures markets (see e.g. Stein, 1961; Hasbrouck, 1995; Gonzalo and Granger, 1995). While findings vary across commodities, most studies find that price discovery occurs in the futures market. There is, however, slightly more evidence of price discovery in the spot market using a fractionally cointegrated model (Dolatabadi et al., 2014, 2015).

A third set of studies estimates the information share across markets for a particular product. An example is Westerlund et al.’s (2014) study that investigates which country leads the price discovery process for crude oil. These authors examine oil prices in the United Kingdom, United States and Oman and find that price discovery occurs in Oman, reflecting the fact that it is the dominant oil producing market.

2.3 Econometric issues in modeling price discovery
The econometric approach to modeling price discovery involves three steps. The first is to ascertain the order of integration of asset prices as a precursor to examining whether there is a long-run relationship between the markets. If two markets are theoretically linked, it follows that any shock in the form of new information will affect both markets. The second step is to establish if prices across the markets are cointegrated. If prices are cointegrated, the question becomes which of the cointegrated markets is the first to absorb the new information. The third step is to
determine in which market price discovery occurs. This is done either through examining the error correction coefficient to ascertain the relative speed of adjustment to the long-run equilibrium price (Gonzalo and Granger, 1995), or estimating the information shares from the error correction model (Hasbrouck, 1995).

2.3.1 Static measures of price discovery

The two main static approaches to measuring price discovery are the information share (IS) method (Hasbrouck 1995) and the component share (CS) method that relies on a permanent-transitory decomposition of the cointegrated system (Gonzalo and Granger, 1995; Harris et al., 1995, 2002). Both approaches are reliant on the vector error correction model (VECM) proposed by Engle and Granger (1987). The approaches differ, however, in that Gonzalo and Granger (1995), and Harris et al. (1995, 2002) focus on the proportion of each market’s innovation that contributes to the common efficient price, while Hasbrouck (1995) focuses on the proportion of the variance in the common efficient price that can be attributed to individual markets. An alternative way of seeing the difference in the approaches is that Gonzalo and Granger (1995), and Harris et al. (1995, 2002) define price discovery in terms of the speed of adjustment to the long-run equilibrium based on the error correction mechanism, while Hasbrouck (1995) defines price discovery in terms of the information share.

2.3.2 Limitations of the CS and IS methods

The main limitation of both the CS and IS methods is the very fact that they are static approaches to price discovery; as such, they are based on market innovations in their reduced form (Scherrer, 2012). However, as price discovery is inherently a dynamic
process in which shocks to prices are likely to be correlated, neither of these static methodologies is able to address the price dynamics involved (Lehmann, 2012). The only way to properly address the price dynamics involved in price discovery is to adopt a structural measure of price discovery in which changes in structural innovations can be appropriately attributed across markets (Scherrer, 2012; Yan and Zivot, 2010).

A second problem is that the CS approach to price discovery is premised on adjustment to the long-run equilibrium price being continuous and linear. For example, this is implicit in the assumption made by Harris et al. (1995, 2002). However, there are several reasons to doubt that adjustment is continuous and linear due to market frictions in the presence of policy interventions and transaction costs. Clearly, market frictions can be explicit or implicit (Gagnon and Karolyi, 2010). Examples of potential explicit market frictions impeding linear adjustment include holding costs (e.g. lending fees, unrealized interest on short-sales proceeds and opportunity cost of capital), short-sale restrictions, taxation and transaction costs (e.g. bid-ask spreads on cross-listed pairs, and foreign exchange and commission) (Chen et al., 2013). Implicit market frictions can be due to cross-border differential adverse selection risks and agency walls between arbitrageur and client (Shleifer and Vishney, 1997). Further, if speculators are active in the spot market-futures market, their trading may destabilize the market and generate a potential non-linear effect of hedging relative to speculation in the futures market’s share of price discovery (Chang et al., 2013).

A third limitation of the CS model relates to the time-varying nature of price discovery. Several studies in related branches such as cointegration show that markets
are characterized by time-varying variance and co-variance. Hence, this time-varying characteristic should have implications for price discovery. What this means is that there may be phases of time over which market X will dominate price discovery, while in other phases it may well be that market Y will dominate price discovery. This will particularly be the case for markets such as oil prices that are prone to volatility (Bessembinder et al., 1995). This is problematic, for example, when examining the price adjustment process in futures markets, in which switches under different market conditions can be expected to induce shifts in the adjustment process to restore the cost-of-carry relationship (Brenner and Kroner, 1995; Caporale et al., 2014).

A fourth limitation with both the CS and IS approaches is that neither allows for the possibility of fractional cointegration. In the context of spot and futures markets, several studies find evidence of fractional integration in the forward premium (see e.g. Baillie and Bollerslev, 1994; Maynard and Phillips, 2001). Hence, fractional integration in the forward premium implies that spot and futures prices are fractionally cointegrated. The reason for this is that although the spot and futures prices themselves may be integrated of order 1 (I(1)), the spread between spot and futures prices will be fractionally integrated of a lower order (Doltabadi et al., 2014).

A fifth limitation of the original CS and IS frameworks is that the microstructure models are limited to two markets. This means that the frameworks do not have the ability to allow for price discovery across multiple markets simultaneously. It also means that, as with any time series measure, that power of the unit root and cointegration tests are limited by the time series and fail to exploit the additional power in a panel.
The final limitation is the econometric challenge in estimating price discovery. The measures are not easy to estimate in practice, however, as they very technical, involve several steps, and are econometrically burdensome. In other words, unlike tests for cointegration which are built into standard software packages and for which many codes are available, the same is not true for price discovery estimators. This thereby limits the application of these tests despite their merits.

2.3.3. Extensions of the CS and IS methods

The CS and IS methods have been extended in various directions to address each of these limitations. First, the fact that the CS and IS frameworks are unable to capture price dynamics has led to the development of structural models of price discovery that allow for changes in structural innovations to be correctly assigned to markets (see e.g. Yan and Zivot, 2010, 2010a; Kim, 2010; Scherrer, 2012). Yan and Zivot (2010) propose the first two-variable structural model of price discovery in which they consider only one common factor. The authors apply their model to the Euro/Yen exchange rate and find that price discovery in this market occurs through the US dollar market. Kim (2010) expands the Yan and Zivot (2010) framework to a three-variable model, sharing two common variables or one cointegrating vector.

Scherrer (2012) presents a structural model of price discovery for a seven-variable model. She first uses the Gonzalo and Granger (1995), and Gonzalo and Ng (2001) method to retrieve permanent and transitory innovations from market residuals in their reduced form. The author then modifies Gonzalo and Ng’s (2001) method to allow the variance of the structural innovations to differ from the identity matrix, which works well with up to three common factors. One advantage of Scherrer’s
(2012) approach compared with those of Kim (2010), and Yan and Zivot (2010) is that the innovations attached to each of the three common factors are allowed to contemporaneously correlate. Another advantage is that Scherrer (2012) decomposes the covariance matrix of the reduced innovations in such a manner that common factors may impact all prices in the long run. The author applies her method to examine price discovery for Brazilian firms cross-listed in Brazil and the United States using high frequency data. She finds that price discovery occurs on the Brazilian stock exchange in the short-run, but that in the long run both exchanges are equally important. Scherrer (2012) presents Monte Carlo evidence suggesting that her structural model outperforms that proposed by Yan and Zivot (2010).

Second, the assumption of continuous and linear adjustment has been relaxed in studies that develop threshold cointegration/error-correction models. Chen et al. (2013) extends the CS method to consider the exchange-respective information shares in Canadian firms cross-listed in Canada and the United States, while Chang et al. (2013) builds on the IS method to examine the effect of non-linearities in the futures market on price discovery in the foreign exchange futures market.

Chen et al. (2013) first use Balke and Fomby’s (1997) threshold cointegration model and then, because dynamics may alternatively be gradual, generalize it to a smooth transition version. Their information share measures are obtained from outer-regime adjustment coefficients based on a two-regime threshold ECM and average coefficient estimates based on a smooth transition ECM. Chen et al.’s (2013) approach has the advantage in that it takes into account non-linearities associated with both abrupt and smooth party changes due to market frictions. Their approach captures the relative contributions to price discovery with a higher degree of precision.
compared with, for example, the ECM in Harris et al. (1995, 2002), which does not address the time-and regime-contingent characteristics of information shocks.

In a related contribution, Chang et al. (2013) examine the effect of the relative size of hedger and speculator open interests on price discovery in the foreign exchange futures market. They first use the IS approach of Hasbrouck (1995) to ascertain the contribution of the EUR-USD and JPY-USD futures markets to the futures/spot price discovery process. With the IS for the futures market, they then estimate a logistic smooth transition model to ascertain how price discovery depends on the trading positions held by hedgers and speculators in the futures markets. Chang et al. (2013) find a non-linear relationship between speculative trading volume and the efficacy of trading volume, in which the effect of speculative trading volume on market efficiency depends on its relation to endogenously determined thresholds.

Third, to accommodate structural change, models have been developed that allow for time varying price dynamics. For example, Caporale et al. (2014) build on the CS approach to take into account parameter instability in the price discovery process. Specifically, in order to detect structural changes in the adjustment coefficients, they estimate time-varying parameter models for the loading weights using a Kalman filter approach. Kalman filtering has the advantage that it does not impose a priori restrictions on the timing of structural breaks, making it useful to ascertain and interpret the effect of any financial turmoil on price discovery. Caporale et al. (2014) apply their model to price discovery in crude oil spot and futures. One interesting conclusion to emerge from their study is that while futures markets play a more important role in price discovery than spot markets, their relative contributions exhibit high instability. Another interesting finding is that spot markets become more
important for price discovery in times of financial crisis such as the Global Financial Crisis.

Fourth, the assumption of a I(0)/I(1) distinction has been relaxed in studies that examine price discovery using FCVAR. Dolatabadi et al. (2014a, 2015) extend Figuerola-Ferretti and Gonzalo’s (2010) model to develop a fractionally cointegrated equilibrium model, with Dolatabadi et al. (2015) using this new model as the theoretical basis on which to apply FCVAR to price discovery in spot and futures markets.

As outlined in Dolatabadi et al. (2014a, 2015), the main advantage of the FCVAR model lies with its flexibility. It permits one to ascertain the cointegrating rank and to jointly estimate the adjustment coefficients and cointegrating relations while accounting for the short-run dynamics. Dolatabadi et al. (2015) find that generally the fractional cointegration parameter is significant, suggesting that a non-fractional model is inappropriate. The authors also find that, when allowing for fractional integration in the long-run equilibrium relations, fewer lags are needed in the autoregressive augmentation of the model, which underpins the utility of the fractional model. Their main substantive finding is that when applying the fractional model, the spot market is slightly more important in the price discovery process relative to the futures market compared with the non-fractional model.

Fifth, the power limitations associated with short time series have been addressed by the development of panel data models of price discovery. De Jong and Schotman (2010) generalize the IS framework to multiple markets. In contrast to the traditional information share that is defined within a reduced-form time series model, De Jong and Schotman (2010) define the information share directly from the unobserved
components model. This approach avoids the wide information share bounds one sometimes gets when using the Hasbrouck (1995) VECM (see e.g. Covrig and Melvin, 2002), particularly in settings in which there are multiple dealers and markets.

More recently, Narayan et al. (2014) propose a panel model in which, conditional on finding panel cointegration, a panel VECM is specified. The panel VECM is used to extract the price discovery coefficients, thus extending the CS and IS frameworks. The conventional panel VECM, such as that proposed in Narayan et al. (2014), has the limitation that if one considers a large number of exchanges (prices) then the model becomes over-parameterized. This comes at the cost of precision in estimation of price discovery. Narayan et al. (2014) only have two price variables and so escape this criticism. However, in most other literature the subject of price discovery involves multiple exchanges, which inhibits the utility of using this approach.

Building on the CS and IS frameworks as well as on De Jong and Schotman (2010), Westerlund et al. (2014) propose a panel data model of prices that have a common factor representation and in which prices are modelled as non-stationary and cointegrated processes. The Westerlund et al. (2014) factor-based panel model is motivated in part by the need to address the econometric complexities with measuring price discovery. Their proposed test is simple and easy to estimate compared to the Generalized Methods of Moments test of De Jong and Schotman (2010), and has two advantages. First, it is free from the limitations of the panel VECM used in Narayan et al. (2014) in that it neither restricts the time-series dimension nor the cross-sectional dimension. Thus, the concern of over-parametrizing the model with the traditional panel VECM is addressed. Second, with existing price discovery tests, there are
restrictions on short-term (transitory) and long-term (permanent) shocks. However, because the Westerlund et al. (2014) test is built on a factor model, it does not impose these restrictions. With the Westerlund et al. (2014) test, all shocks are entertained in their original form, that is, without any restriction.

2.4 State of the art

The above discussion suggests that by using the CS and IS frameworks as its starting point the price discovery literature has followed two directions. One involves developing panel models of price discovery, while the other focuses on relaxing various assumptions of the standard IS and CS frameworks in a time series context.

Perhaps the nearest to the state-of-the-art in the traditional IS/CS framework is the seminal ‘unifying’ contribution of Figuerola-Ferretti and Gonzalo (2010), who demonstrate that the Garbade and Silber (1983) framework coincides exactly with the permanent component in the Gonzalo and Granger (1995) permanent-transitory decomposition. This has paved the way for treating the CVAR model as the natural model to analyze price discovery, and provided a relatively simple way of ascertaining which of two prices is dominant in terms of long-run price discovery.

Figuerola-Ferretti and Gonzalo’s (2010) framework, however, no longer represents the state-of-the-art in price discovery per se. Other models have been developed to address the limitations of the CS/IS framework, with Figuerola-Ferretti and Gonzalo’s (2010) framework now serving as a platform for the FCVAR proposed by Dolatabadi et al. (2015, in press) Further, the CS and IS frameworks have been extended in other directions, including the development of structural models to better model price dynamics, models that take into account non-linearities or structural change, as well
as panel models to deal with multiple markets and/or small time series. Because these contributions emanate from myriad directions, it is difficult at this point to pinpoint one state-of-the-art method for determining price discovery. The preferred model really comes down to the issue affecting price discovery that one wants to focus on. Perhaps the most that can be said is that these developments are at the frontier and in compact represent the state-of-the-art in modeling price discovery.

2.5. Suggestions for future research

Given that the time series literature has moved in two directions, namely, the development of panel models and relaxing assumptions underpinning the CS and IS frameworks, one avenue for future research could be to integrate some of the developments of these two avenues of inquiry into a panel framework. Obvious examples would be to incorporate non-linearities, structural change or fractionally integrated processes into recently proposed panel models.

A second avenue for future research could be to further develop structural models of price discovery. The most general model of this form in the literature appears to be from Scherrer (2012), who presents a method to measure the instantaneous effects of permanent shocks on prices. Her model allows the observed price to depend on three common factors, which are allowed to be contemporaneously correlated, providing the required conditions for linkages across common factors. One potential extension to the Scherrer (2012) framework could be to extend the number of common factors.

A third direction for future research could be to develop regime switching price discovery models, which may be sufficient to allow an examination of price discovery during economic expansions and recessions about which nothing is known.
A fourth direction for future research in this area could follow on from recent developments in econometric modelling rather than extending its frontier. Recent innovations in econometric modelling have challenged long-held beliefs about the markets in which price discovery occurs. One example is that incorporating fractional cointegration seems to provide more evidence that price discovery occurs in the spot market. Another example is that allowing for the effects of financial crises suggests that incorporating time-varying dynamics provides evidence that fundamental factors may have a greater role in driving prices, particularly in periods of financial crises. Future research could address why one finds more evidence of price discovery in specific spot markets with fractional integration or time-varying dynamics given the economic fundamentals in particular markets.

3. Price Predictability

3.1. Motivation and implications

Two variables that commonly appear in the financial econometrics literature on predictability are securities and exchange rates. Prior to the seminal review by Fama (1970), empirical studies provided strong evidence that securities’ markets are predictable using past returns. Fama (1970) defines this form of return predictability as the weak form of the efficient market hypothesis (EMH). However, relying only on past returns lacks power as past realised returns are noisy measures of expected returns (Fama, 1991). To improve the power of return predictability tests, predictor variables that are less noisy proxies for expected returns could be used (Fama, 1991). Some early evidence that short-horizon returns can be predicted from past values of other financial and macroeconomic variables include expected inflation (Bodie, 1976; Fama, 1981), short-term interest rates (Fama and Schwert, 1977), dividend yields
(Rozeff, 1984; Shiller, 1984; Fama and French, 1988), and earnings to price ratios (Campbell and Shiller, 1988), all of which are able to explain a larger fraction of the return variation for longer return horizons. In the past three decades an influx of other predictors of stock markets has appeared in the literature, including the seminal work of Campbell and Shiller (1988) that shows that market returns can be predicted to some extent by using financial ratios such as dividend yields and price-earnings ratios (see Fama, 1991, and Malkiel, 2004, for a discussion of the success of past studies in predicting future equity returns).

The success of predictability studies in the 1980s and 1990s led many studies to question the EMH, although Fama (1991) suggests that the results can be achieved within the confinement of rationality, thus extending the definition of the weak-form EMH to include past movements in variables other than past returns (see also Lim and Brooks, 2009). However, many studies, particularly in the behavioural finance literature, conclude that investors are prone to irrational behavior (or psychological pitfalls); hence, this implies the potential to make abnormal profits (see Malkiel, 2003; Schwert, 2003; Malkiel et al., 2005; Lo, 2004; Yen and Lee, 2008). This change in attitude towards stock market predictability has been fuelled by the growth of computer power and has led to the development of a number of estimation and predictability techniques.

Thus, while testing exchange rate predictability is strongly supported by economic theory, the challenge since Meese and Rogoff (1983) has largely remained one of finding predictive models for the exchange rate beyond the simple random walk model.
The rest of this part of the survey is concerned with financial econometric developments in the predictability literature on securities and exchange rates – by far the most common financial variables covered in this literature. Our focus is on the predictability of these two variables using past values of financial and macroeconomic variables. Although this literature is vast, our particular focus is on advances in econometric techniques. We initially provide an overview of predictive modelling using unrestricted and restricted vector autoregressions (VARs), including both panel and time series analyses. This is followed by a review of recent developments in financial econometries in forecasting financial variables.

3.2. Predictive modelling using unrestricted and restricted VARs
Sims (1980) highlights the importance of the VAR formulation as a convenient representation for forecasting and estimation, while Engle and Granger (1987) establish that an error correction framework provides the means for testing, estimating and forecasting a cointegrated system (see also Engle et al., 1989; Christoffersen and Diebold, 1998). Engle and Yoo (1987) illustrate that when the unrestricted VAR (in levels) is compared to cointegrated systems, it leads to suboptimal forecasts at the longer horizon. Hence, they conclude that it is important to “build models designed for cointegrated time series or vector autoregressive error-correction models (VECMs)” (p. 159) (see also Phillips, 1998; Hoffman and Rasche, 1996; Clement and Hendry, 1995).

The advancement of this technique as a forecasting tool has also gone hand-in-hand with the quest to explain foreign exchange rate behavior within the confines of the monetary model of exchange rate. Indeed, a long-run cointegrating relationship
between the exchange rate and monetary fundamentals is a central tenet of the monetary model of the exchange rate. Thus, if this model depicts the long-run behavior of the exchange rate correctly, then the exchange rate only deviates from monetary fundamentals in the short-run, making the vector error correction model (VECM) in forecasting exchange rates a natural contender (see also Groen, 2005). While early studies such as Mark (1995), Chin and Meese (1995), and Groen (2000) show the importance of first establishing the cointegration relationship between the monetary fundamentals and exchange rate in the exchange rate predictability exercise, predictability in these studies has been confined only to the long horizon. These long horizon-to-weak-to-no predictability results, however, have fueled a search for a more versatile, efficient and powerful error-correction framework predictability test.

In terms of versatility then, and unlike Engle and Granger’s (1987) test, Johansen (1988, 1991) allows for more than one cointegrating relationship. The autoregressive distributed lag (ARDL) - bounds testing methodology of Pesaran and Shin (1999), and Pesaran et al. (2001) allows for a mixture of I(0) and I(1) variables, provides for more variables to enter the long run model, and has been proven useful in predicting long-run asset returns. The fractional cointegration test developed by Johansen (2008) is suitable for several financial variables well described by long-memory fractionally integrated processes, and has also been proven useful in explaining the return-volatility relationship (see e.g. Bollerslev et al., 2013). With evidence of parameter instability in the exchange rate as well as in the stock market, this time-varying cointegration approach represents a logical tool. Further, Park and Park (2013) show that the time-varying approach has strong predictive power for future changes in

---

1 Hence, predictability of exchange rates and econometric developments in the area of predictability using the VECM framework have gone hand-in-hand.
exchange rates through in-sample and out-of-sample analysis.

In order to address the low power of time series, the Engle-Granger type tests have also followed developments in testing for a unit root. The predictability literature has also analyzed time series across similar cross-sections in panel datasets. As Westerlund and Narayan (2014c, p. 2) point out, “…the use of panel rather than time series data not only increases the total number of observations and their variation, but also reduces the noise coming from the individual time series regressions. This is reflected in the power of the resulting panel predictability test, which is increasing in both $N$ (number of cross-sections) and $T$ (time span), as opposed to a time series/unit-by-unit approach where power is only increasing in $T$. Thus, from a power/precision point of view, a joint (panel) approach is always preferred. Furthermore, since power is increasing in both $N$ and $T$, this means that in panels one can effectively compensate for a relatively small $T$ by having a relatively large $N$, and vice versa”.

As indicated above, the panel version of the Engle and Granger (1987) test involves estimating the panel long-run model with a panel unit root test (see e.g. Levin et al., 2002; Im et al., 2003) being conducted on the residuals. This approach implies both homogeneous long-run coefficients and homogenous adjustment parameters, with only the serial correlation in the residuals panel unit root test assumed to be heterogeneous (Groen, 2000). Error-correction based forecasting models from these techniques have the disadvantage that the differences in the speed of adjustment and dynamics across different countries are not taken into account (see e.g. Groen, 2000). Commonly used variants of the panel-based Engle-Granger approach are the Pedroni (1995), and Kao (1997) tests, which assume heterogeneity in the model parameter and
statistics. These tests are based on cross-sectional averages of the individual parameters and statistics. However, because these parameters and statistics are independent, Groen and Kleibergen (2003) argue that these tests do not capture the panel dimension that is crucial to enhance the power of cointegration tests. Thus, to allow for interdependence between panels, Groen and Kleibergen (2003) stack a fixed number of VECM of different individuals into a joint panel VECM and develop a likelihood-based framework for cointegration analysis, generalising Johansen (1988, 1991) (see also Larsson et al., 2001).

These aforementioned tests have become increasingly efficient over time as they incorporate better estimating procedures for forecasting parameters. For example, apart from using ordinary least squares (OLS) as the standard estimation method, methods such as fully-modified OLS (FMOLS) and dynamic ordinary least squares (DOLS) (see Saikkonen, 1991; Stock and Watson, 1993; Kao and Chiang, 2000) have been developed for time series and panels, with Kao and Chiang’s (2000) examination of the finite-sample properties of the DOLS estimator finding it to be superior to both OLS and FMOLS in panel regressions.

In panel-based forecasting analysis, it has become common to pool the parameters of the forecasting equation, as in Mark and Sul (2001). Using simulations, Westerlund and Basher (2007) show that pooling individual prediction tests (such as the Theil $U$ statistic and the $S$ prediction statistics of Diebold and Mariano, 1995) can lead to substantial power gains. These authors further show that pooling only the parameters of the forecasting equation does not seem to generate more powerful tests.
3.3. Developments in econometric modelling of forecasting financial variables
Beginning with Mankiw and Shapiro (1986), and Stambaugh (1986), innovation in financial econometrics in the predictability literature over the past three decades has primarily addressed various challenges which we characterize under three broad topics. These include the salient features of the predictive models and its predictors, predictability in the presence of model instability and uncertainty, and combination forecasts. We focus on each of these three areas in turn.

3.3.1 Salient features of predictive models
(a) Persistent regressors
There is now strong evidence that widely accepted predictors of excess returns (such as dividend-price ratio, earnings price ratio and measures of interest rates) are persistent and behave like non-stationary variables, although they may not be fully integrated (see Campbell and Yogo, 2006; Lewellen, 2004; Torous et al., 2004). Meanwhile, the dependent variable (e.g. excess returns) may exhibit less persistence than their predictors (Maynard et al., 2013). Cavanagh et al. (1995) show that classifying the predictor variables as I(0) or I(1) leads to large size distortions when regressors have local-to-unit roots. Predictability studies indicate that ignoring this persistence in the regressors can lead to over-rejection of the null hypothesis using conventional critical values (Campbell and Yogo, 2006; Torous et al., 2004). Furthermore, Phillips and Magdalinos (2007a, b), and Magdalinos and Phillips (2009a) recently analysed a wide range of close-to-unity processes and named those in the vicinity of stationary, but not quite stationary, as mildly integrated processes. Many variables, including Treasury Bill rates, show varying degrees of persistence and seem to be mildly integrated processes.
While Phillip and Lees (2013) only recently accounted for the mildly integrated process in predictive regression methodology, there are several studies in the predictability literature that have paid attention to nearly integrated variables. Elliot and Stock (1994), for instance, propose an asymptotic framework in which the regressor is allowed to follow an autoregressive process with a root near to unity so that it approximates a finite sample distribution of test statistics when the predictor is persistent. However, allowing for persistence in the regressor leads to a nonstandard limit distribution of the t-statistic that becomes dependent on a nuisance parameter, and this cannot be consistently estimated (see Cavanagh et al., 1995; Torous et al., 2004; Campbell and Yogo, 2006). Torous et al. (2004), and Campbell and Yogo (2006) are other studies that develop asymptotics for inference using a local-to-unity autoregressive specification for a single regressor. The Campbell and Yogo (2006) method has been particularly popular among empirical researchers, is an industry standard in finance, and provides a benchmark for competitors (Phillips, 2014, p. 1191).

These aforementioned studies follow Cavanagh et al. (1995), and use Bonferroni intervals of the coefficient (of the predictor variables) in which the dependence from the localising coefficient, \( c \), is eliminated. Nonetheless, Bonferroni-type modelling requires simulations to compute critical values to perform inference and confidence interval construction, since the limit distribution employed in the calculations is nonstandard (Phillips and Lee, 2013). These tests are also difficult to extend beyond one regressor, which is not a desirable feature as there are typically several predictors for each financial variable (Phillips and Lee, 2013). Furthermore, empirical size may be substantially lower than nominal size and result in a conservative test. Cavanagh et
al. (1995) suggest aligning the nominal size to a desired level asymptotically (see Torous et al., 2004), while Phillips and Less (2013) argue that the power of the conservative test is often negligible in near local alternatives to the null of non-predictability.

To address the uncertainty surrounding the root, these papers follow Stock (1991) to construct confidence intervals for the largest autoregressive root of many explanatory variables used in predictive regressions. However, Phillips (2014) shows that Stock’s confidence intervals for the root are invalid and seriously bias asymptotically when $c \rightarrow -\infty$. This failure in the approach leads to poor performance in predictive regression tests based on Bonferroni methods when the regressor is stationary or mildly integrated (Phillips, 2014). Elliott et al. (2012) develop a ‘nearly optimal’ test so that the procedure does not suffer from the problems inherent in the Campbell–Yogo test. Furthermore, as Phillips (2014) argues, the computational and design considerations make this method most useful in the single regressor case.

Finally, in these local-to-unity cases, the finite sample bias in estimation (discussed below) is still present in the limit, and correcting the asymptotic bias is generally not possible since the bias depends on the localizing coefficient, $c$, and this parameter is not consistently estimable (Phillips and Lee, 2013). Lewellen (2004) extends the work of Stambaugh (1999) to account for the finite sample bias and persistence in the regressor. Lewellen uses conditional tests (which are estimated from OLS) to show that predictability can be improved if information in the predictor autocorrelation, especially when it is close to one, is utilized. Other approaches also provide solutions to the nonstandard and non-pivotal limit distribution problem, including the
conditional likelihood approach with sufficient statistics developed by Jansson and Moreira (2006), a control function approach by Elliott (2011), and the instrumental variable (IVX) approach proposed by Magdalinos and Phillips (2009b) (see Phillips and Lee, 2013, for a detailed discussion of each of these approaches).

In particular, Phillips and Lee (2013) extend Magdalinos and Phillips (2009b) IVX approach to solve most existing problems in predictive regression methods. The IVX approach generates a less persistent instrument for the regressor (X) than the regressor itself, which uses no extraneous information so that the instrument relies only on the regressor. Phillips and Lee (2013) add to this by allowing for multivariate regressors, including predictive processes whose roots may lie in a wide vicinity of unity and mildly integrated, local-to-unity and unit root regressors, all of which are examined with standard chi-square tests and can be applied easily to long-horizon predictive regressions. While the implementation of this approach is by straightforward linear regression, the method requires user input on the parameter choice for instrument construction (Phillips, 2014).

The above studies address size distortions in the context of the local-to-unity case. However, persistence can also manifest itself in the form of long memory (Maynard et al., 2013). Fractionally differenced processes can be observed in several predictors such as forward premium (Baillie and Bollerslev, 1994; Maynard and Phillips, 2001), volatility (Comte and Renault, 1998; Baillie and Bollerslev, 2000) and dividend yields (Koustas and Serletis, 2005).
Maynard et al. (2013) develop a predictive test based on a two-stage rebalancing procedure in which a semiparametric or parametric estimator may be used to estimate the degree of long memory in the regressor in the first stage. In the second stage, the regression is rebalanced by fractionally differencing the regressor. This also rebalances the alternative hypothesis while leaving the null hypothesis unchanged and thus allowing for a valid test of non-predictability. The authors’ Monte Carlo study confirms that fractionally differencing the regressor removes the source of the size distortion and yields a t statistic in the second-stage regression with correct size. They also find that estimation and inference in the second stage are robust to estimation error or even modest misspecification in the first stage.

The latter result has connections with recent studies that show that the functional form may affect the power of predictive tests under non-stationarity with a parametric approach (Kasparis et al., 2014). And this also may be true for some nonparametric nonstationary specifications (Wang and Phillips, 2009). With this in mind and using the recent work of Wang and Phillips (2009), Kasparis et al. (2014), in particular, develop a unifying framework for inference in predictive regressions where the predictor has unknown integration properties. They propose two easily implemented nonparametric F-tests in which limit distributions is nuisance parameter free and holds for a wide range of predictors, including stationary as well as non-stationary fractional and near unit root processes. Kasparis et al.’s simulations suggest that this test is more powerful than existing parametric tests (e.g., Jansson and Moreira, 2006) when deviations from unity are large or the predictive regression is nonlinear. Karemara and Kim (2006) suggest that using long memory models for forecasting nominal exchange rates is a viable alternative to conventional models. They show that
an autoregressive fractionally integrated moving-average (ARFIMA) model is more efficient than the random walk model in steps-ahead forecasts beyond one month. The authors also find it to be more efficient than the random walk model in multi-step-ahead forecasts.

(b) Endogeneity

In financial predictive models where returns are regressed on a lagged predictor variable (whether for exchange rates or stocks), the regression results will suffer from small sample bias if the regression disturbances are correlated with the predictor’s innovations. This creates the so-called embedded endogeneity problem that leads to biased estimates. Stambaugh (1999) illustrates through the example of excess stock returns and dividend yield that in the presence of this bias, “…finite-sample estimation and inference become less straightforward, for at least two reasons. First, the ordinary least squares (OLS) estimators, although consistent, are biased and have sampling distributions that differ from those in the standard setting. Second, differences between classical and Bayesian methods become more apparent in the presence of this Stambaugh bias, whereas those approaches are distinguished less often in the standard regression setting”. Stambaugh’s (1999) approach to accounting for endogeneity applies in the case of a single predictor, while more recently, Amihud and Hurvich (2004), and Amihud et al. (2009) account for endogeneity in the presence of multiple predictors. Furthermore, Stambaugh accounts for endogeneity assuming that the predictive regressor’s stochastic explanatory variable is stationary. Lewellen (2004) allows for a persistent regressor and Stambaugh’s bias in his study, and shows that without accounting for persistence the corrections for small sample biases can substantially understate forecasting power.
Similarly, Cai and Wang (2014) are concerned with the presence of endogeneity in the model and persistence of the regressor, although their approach is different from Lewellen (2004) because they work with multiple predictors. Cai and Wang (2014) address endogeneity by adopting the Amihud and Hurvich (2004), and Amihud et al. (2009) linear projection method, and then using a two-step estimation procedure to manage both highly persistent and non-stationary predictors.

(c) Heteroskedasticity

Another important feature of financial predictive regression is the presence of noise or time-varying ARCH/GARCH effects in asset returns. The long horizon predictive regressions as in Torous et al. (2004 in which the predictor variable may be monthly but the returns variable can be measured over one, two, or five year/s, have long been used as a way to reduce noise in stock returns (Campbell and Yogo, 2006). This is because, under the alternative hypothesis that returns are predictable, the variance of the return increases less than proportionally with the investment horizon (Campbell and Yogo, 2006). Compared to the long horizon tests, the procedures proposed in Lewellen (2004), and Campbell and Yogo (2006) are even more powerful as they reduce noise not only under the alternative hypothesis, but also under the null.

Westerlund and Narayan’s (2014b) study proposes a test that exploits the information contained in the heteroskedasticity of returns, which is found to lead to higher power. Similar to the above studies, their generalised least squares (GLS) test accounts for persistence in regressors and endogeneity issues. The asymptotic test’s dependence on nuisance parameters reflecting persistence, endogeneity and heteroskedasticity in the data, however, means that the test can suffer from size distortions. To account for this
lack of robustness, the authors propose a subsample feasible quasi-GLS (FQGLS) test that has correct asymptotic size even in the presence of nuisance parameters. This differs from many of the above tests, which use Bonferroni-type tests that are known to be conservative.

3.3.2 Predictability in the presence of model instability and uncertainty

This literature is vast and complex. In this section, we focus only on areas that have seen recent econometric developments or are seen as emerging areas. We begin by looking at model instability, focusing on studies that have developed the predictability framework under the assumption of structural break uncertainty (see e.g. Pesaran and Timmermann, 1995, 2002). We also review another group of studies that treat structural breaks as being deterministic in nature and, hence, address issues surrounding the number and duration of breaks. Related to this is the idea that there is a stochastic process underlying the structural breaks that is used to predict future structural breaks. The time-varying literature takes the issue of model instability to the extreme, and is premised on the notion that there are not just one, two, or five structural breaks, but many structural breaks across each period that render parameter uncertainty. To this end, this literature argues for time-varying approaches for estimating parameters, and developing many tests to both detect the phenomenon and to develop such approaches. Furthermore, choosing the right set of variables for forecasting has led to some studies trying to deal with model uncertainty. Finally, there are those we refer to as ‘hybrid studies’ within the literature on model instability and uncertainty that address aspects of instability and uncertainty within the same predictability framework.
3.3.3 Model instability

(a) Structural break uncertainty

Structural breaks refer to both permanent shifts and ‘breaks’ (temporary jumps) in the parameters of the return generating process. There could be various reasons for permanent shifts, including legislative, institutional or technological changes, shifts in economic policy, or even large macroeconomic shocks such as the doubling or quadrupling of oil prices that have been experienced over recent decades (Pesaran et al., 2006). Breaks (or jumps) in the parameters may also arise due to a number of factors, including major changes in market sentiment, the creation or bursting of speculative bubbles, or regime switches that affect monetary and debt management policies. The latter, for example, might include switches from money supply targeting to inflation targeting, or from short-term to long-term debt instruments (Pesaran and Timmermann, 2002).

To address parameter instability, many studies are concerned with the problem of determining in real time how much historical information to use when estimating a forecasting model. Studies use a range of methods, such as a rolling window of observations with a fixed size to generate the forecast, discounted least squares that assigns smaller weights to observations further away from the point of the prediction, or recursive least squares that allows for expansion of the observation window to allow for new information (see Pesaran and Timmermann, 2002). While Pesaran and Timmermann (1995) allow the forecasting model to change over time, they do mention in their 2002 paper that this may be inappropriate in the presence of breaks in the parameters of the forecasting model. The choice of estimation window of a forecasting model is endogenised in Cooper and Gulen (1999), but that study is
constrained to either rolling windows of fixed length or an expanding window. However, none of these methods explicitly detect, and condition on, the occurrence of one or several breaks. This is a concern of Pesaran and Timmermann (2002), who develop a conditionally time-varying window size using a two-stage approach. A reversed ordered Cusum (ROC) test is developed to detect the most recent breakpoint. This test is related to the standard CUSUM test developed by Brown et al. (1975) (see Tian and Anderson, 2014, for a detailed discussion). When adopted recursively through time, the reversed ordered Cusum procedure yields a sequence of estimation windows whose lengths effectively indicate the ‘memory’ of the return model under consideration. The resulting sequence of ROC tests statistics is used to choose and forecast from a single post-break estimation window.

(b) Deterministic structural breaks

There are studies that treat structural breaks as being deterministic in nature. The aim of these studies is to explicitly identify the number and duration of break dates, and to examine the influence of the structural breaks in the forecasting exercise. As implied above, the identification of the number and duration of the structural breaks or regime shifts is critical here. Early studies determined the number and duration of structural breaks exogenously. However, given the uncertainty with this approach, various studies have attempted to develop techniques that allow determination of the breaks endogenously (see e.g. Rapach and Wohar, 2006).

The Chow (1960) procedure is a familiar test with the null hypothesis of no structural change, but one may use the standard F-test to test this against the alternative only if the breakpoint is known. Quandt (1960) suggests using the maximum F-statistic over all values of break dates as a means to test parameter stability. This search over a set
of dependent F-statistics affects the asymptotic distribution of the test, which ceases to be χ² (Elliott & Müller, 2003). Expanding on Quandt (1960), Andrews (1993) derives the limiting distribution of the supremum of the Fₖ statistics (SupF statistic), which is nonstandard and dependent on a trimming parameter. At any given trimming parameter, the null hypothesis of no structural break can be tested using the asymptotic critical values in Andrews (1993) and, if the null is rejected, the break date can be estimated (see also Rapach and Wohar, 2006). Hansen (2000) develops a heteroskedastic fixed-regressor bootstrap procedure that delivers the correct asymptotic distribution for the SupF statistic in the presence of general nonstationarities in the regressors, including mean and variance breaks and unit roots, which Andrew’s asymptotic distribution fails to account for. Bai (1997), and Bai and Perron (1998, 2003a, 2004) explicitly test for multiple structural breaks at unknown dates in the bivariate predictive regression models. Apart from allowing for multiple breaks, the Bai and Perron methodology accounts for autocorrelation in the regression model residuals, heteroskedasticity in the residuals, and different moment matrices for the regressors in the different regimes when computing test statistics and confidence intervals for the break dates and regression coefficients.

(c) Stochastic structural breaks in forecasting

Pesaran et al. (2006) develop a test to examine how future values of the variables of interest might be affected by multiple breaks. As the authors explain, this question cannot treat structural breaks as deterministic (as in the case of past structural breaks) unless they are known; rather, it requires a modelling of the stochastic process underlying the breaks. They propose a Bayesian estimation and prediction procedure that allows for the possibility of new breaks occurring over the forecast horizon, taking into account the size and duration of past breaks (if any) by means of a
hierarchical hidden Markov chain model. Here, predictions form by integrating parameters from the meta-distribution that characterizes the stochastic break-point process.

(d) Time-varying parameter - parameter uncertainty
In the stock market predictability literature, various studies, including Paye and Timmermann (2006), and Goyal and Welch (2008), show that the magnitude of asset return predictability is distinctly time-varying and unstable; however, this is not just limited to stock returns but is true of a wide range of time series (see Stock and Watson, 1996). Merton’s (1973) intertemporal CAPM also suggests that time-varying risk aversion can lead to a time-varying relationship between stock returns and predictive variables (see Zhu and Zhu, 2013). And if predictability of returns partly reflects market inefficiencies and not just time-varying risk premia, then predictive relationships should disappear once discovered, provided that sufficient capital is allocated towards exploiting them (Pesaran and Timmermann, 2002). Several studies show the importance of business cycles in dealing with predictability in returns (see e.g. Fama and French, 1989; Lettau and Ludvigson, 2001; Cooper and Priestly, 2009; Henkel et al., 2011; Nitschka, 2013, 2014). Collectively, these studies make the point that parameter uncertainty or the time varying nature of parameters seem to be driven by business cycles.

While the studies on structural breaks/jump work mainly with one to five breaks, the time-varying parameter literature is concerned with large as well as small parameter changes that occur now and then. Furthermore, because the breaking process can occur in any number of ways – for instance, over the review period, breaks can occur
every year or every second year, in clusters, or randomly – several studies have focused on different breaking processes. To give a flavour, studies such as Shively (1988b) work with Gaussian breaks of constant variance every period, Andrea and Ploberger (1994) use a finite number of breaks by employing an average asymptotic, and Bai and Perron (1998, 2003) work with the maximum of a sequence of F-statistics for their power criterion. To compare various breaking processes used in the model instability and time-varying parameter literature, Elliott and Müller (2003), in particular, show that leaving the exact breaking process unspecified (apart from a scaling parameter) does not result in a loss of power, at least asymptotically.

Here, as explained by Hackl and Westlund (1989), time-varying studies are concerned with tests to detect non-constant parameters in regressions and, if the null of constant regression is rejected, implement time varying approaches. A number of studies examine the hypothesis of constant regression against the alternative hypothesis that the parameter is stochastic and follows a first order autoregressive process – in this case, the non-constant parameter is modelled to deviate only temporarily from zero, so that the ‘long-run’ value remains the estimated parameter (see Shively, 1988a). Most Markov switching models with recurring states and threshold autoregressive models are also closely related to the case of a stable autoregressive process for the non-constant parameter (see also Elliott and Müller, 2003). Most studies in this area, however, consider models in which deviations of the non-constant parameter from zero are permanent. In these models the alternative hypothesis is that the non-constant parameter follows a random walk. For the ‘unobserved components’ model, see Chernoff and Zacks (1964), and Nyblom and Mäkeläinen (1983). For more general

(e) Model uncertainty

Similarly, while some predictive variables such as financial ratios, including dividend yields and price-earnings ratios, are popular in the predictability literature, many other predictability variables are also covered. Given the high volume of predictors, it might often seem unclear to an investor as to which set of variables comprises the correct predictive variables (see also Schrimpf, 2010). Hence, an important question that arises is how investors choose between so many competing models (Pesaran and Timmermann, 2002). Moreover, choosing the best model or model thinning is restrictive, as information from the discarded models is ignored in each period (Aiolfi and Favero, 2005). Beginning with Cremers (2002), and Avramov (2002), studies examine ways to efficiently deal with this model uncertainty. As Aiolfi and Favero (2005) explain, a natural way to interpret model uncertainty is to refrain from the assumption of the existence of a ‘true’ model and instead attach probabilities to different possible models – an approach known as ‘Bayesian model averaging’. Cremers (2002), and Avramov (2002) use a pure Bayesian model averaging approach that requires prior elicitation for the relevant parameters, conditional on the different models. Avramov (2002) applies an empirical Bayesian approach for prior elicitation, which uses data-information from the sample for prior elicitation. However, such an

---

2 Among others, predictors of equities in the literature include expected inflation (Bodie, 1976; Fama, 1981), short-term interest rates (Fama and Schwert, 1977; Hodrick, 1992), output gap (Cooper and Priestly, 2009; Møller et al., 2014); the consumption wealth ratio (Lettau and Ludvigson, 2001); consumer confidence (Møller et al., 2014); and investor sentiments/attention/anchoring behaviours (Li and Yu, 2012). The fact that many variables have been found to be valuable predictors of returns has raised concern that apparent predictability may be a result of data-snooping rather than of genuine variation in economic risk-premia (see e.g. Goyal and Welch, 2008). Authors such as Schrimpf (2010), Cremers (2002), and Avramov (2002) regard the long list of apparent predictive variables as a sign of model uncertainty.
approach has been criticized for using information about the dependent variable, which violates the rules of probability necessary for conditioning (Fernández et al., 2001). In the Bayesian tradition, Cremers (2002) specifies subjective prior distributions that are based on different beliefs about predictability. The specification of prior beliefs, however, can be a problematic task when the set of models becomes very large (Schrimpf, 2010). To avoid the drawback of the dependence on prior distribution, studies such as Schrimpf (2010) adopt the Salai-Martin et al. (2004) approach, that is, the Bayesian averaging of classical estimates. Furthermore, a pure Bayesian model averaging approach takes all predictive variables as exogenous. To remedy this, Schrimpf (2010) combines the Bayesian feature of model averaging with coefficients estimated by standard OLS, which are adjusted for endogeneity bias following Amihud and Hurvich (2004), and Amihud et al. (2009).

3.3.4 Combination forecasts
The pioneering work of Bates and Granger (1969) suggests that if a number of unbiased forecasts of the same future variable are available, then, rather than using one, the forecasts can always be combined in such a way that the composite forecast has variance less than or equal to any of the competing forecasts. Since Bates and Granger’s (1969) work, various ‘model averaging’ studies have tested the efficacy of combination forecasts from different models (similar to Bates and Granger, 1969). While there is clear support for combination forecasts in this literature, it is not obvious which combination scheme is the best. Bates and Granger (1969) find that combination forecasts that allow the weights to change can lead to better forecasts than those that result from the application of a constant weight determined after noting all individual forecast errors. Makridakis and Winkler (1983) show that
differential weights that are proportional to the reciprocal of a sum of squared errors work better than a simple average.

Recent studies have resorted to combining models using either the same or different estimation procedure but different predictors. Examples include shrinkage methods such as ridge regression (Hoerl and Kennard, 1970), bootstrap aggregation (or bagging) (Breiman, 1996), and the least absolute shrinkage and selection operator (Lasso) (Tibshirani, 1996). These studies apply various flexible weighting to individual predictors and include: bagging (Breiman, 1996), which applies differential shrinkage weights to each coefficient; adaptive Lasso (Zou, 2006), which applies variable-specific weights to the individual predictors in a data-dependent adaptive manner; the elastic net (Zou and Hastie, 2005; Zou and Zhang, 2009), which introduces extra parameters to control for the penalty of including additional variables; and Bayesian methods such as adaptive Monte Carlo (Lamnisos et al., 2012; Tian and Anderson, 2014), which assign heavier weights to forecasts that use more recent information on different estimation windows. Elliot et al. (2013) show that combinations of subset regressions can produce more accurate forecasts than conventional approaches based on equal-weighted forecasts, combinations of univariate forecasts, or forecasts generated by methods such as bagging, ridge regression or Bayesian model averaging.

While combination forecasting is becoming popular in the economics literature, predictability literature in finance is only just beginning to catch up (see Rapach et al., 2010). Only a few methods have been developed/applied in the finance literature. For instance, a fixed number of predictors of stock returns are combined using an equal-
weighted scheme on the basis of complete subset regressions (Elliot et al., 2013), averaging of OLS and Kalman filter of Mamaysky et al. (2008) models of mutual funds (Mamaysky et al., 2007), equal-weighted methods to combine many univariate equity premium models (Rapach et al., 2010), and equal-weighted combinations of forecasts of stock returns from all possible two variable models (Aiolfi and Favero, 2003).

Overall, the strong message in recent studies is that combination forecasting is better than standalone forecasting. And while it is unclear which combination method is the best, a scheme that applies differential weights seems to work well in this literature. Moreover, it appears using a few schemes is the best approach.

3.4 State of the art

3.4.1 Persistence

At present, Bonferroni-type procedures are regarded as the state-of-the-art in this literature, but they do have some properties that limit their adoption in empirical work (Phillip and Lee, 2013). Other alternatives are available, including Phillip and Lee’s (2013) parametric test and Kasparis et al.’s (2014) non-parametric test. Both tests can be easily applied and are versatile enough to accommodate a wide range of predictors, including stationary as well as non-stationary fractional and near unit root processes.

3.4.2 Persistence, endogeneity and heterogeneity

Modeling approaches that take account of endogeneity, heterogeneity and persistence and can be implemented easily are state-of-the-art. Westerlund and Narayan (2014b) (one predictor) and Lewellen (2004) (multiple predictors) are two approaches for time series analysis, while Westerlund and Narayan (2014a) present a panel-based test.
3.4.3 Model instability and uncertainty

There are several standalone developments in this literature that have made significant contributions. Methods that will make the most difference include those that incorporate several features of model instability and uncertainty. For instance, using Bayesian model averaging techniques, Pettenuzzo and Timmermann (2011) account for model uncertainty and instability; that is, investors are not assumed to know the true model or its parameter values, nor are they assumed to know the number, timing or magnitude of past or future breaks. Instead, they come with prior beliefs about the meta-distribution from which current and future values of the parameters of the return model are drawn and update these beliefs efficiently as new data are observed.

Zhu and Zhu (2013), on the other hand, develop a Bayesian regime-switching combination (BRSC) approach that explicitly incorporates model, regime and parameter uncertainty to examine the out-of-sample return forecasting problem. Here they incorporate model uncertainty, as in Barberis (2000), and Cremers (2002). The uncertainty about parameters is summarized by the posterior distribution of the parameters given by the data. Instead of constructing the distribution of expected returns conditional on fixed parameter estimates, the Bayesian method integrates over the uncertainty in the parameters captured by the posterior distribution. To account for time varying predictability and address the regime uncertainty issue, the regime-switching combination approach is used (see Rapach et al., 2010) to account for model uncertainty and efficiently use the information contained in a large collection of candidate predictive variables.
3.4.4 Combination forecasts

With many models and predictors of securities and exchange rates, this is one important area of research. However, it is difficult to suggest state-of-the-art techniques here, as many developments in this area have occurred outside the financial econometrics literature that currently relies heavily on equal-weighted forecasts from two to many univariate models of returns or equity premium or model averaging schemes. A recent paper by Elliot et al., (2013) shows that combinations of subset regressions can produce more accurate forecasts of stock returns than conventional approaches based on equal-weighted forecasts, combinations of univariate forecasts, or forecasts generated by methods such as bagging, ridge regression or Bayesian model averaging.

3.5 Suggestions for future research

This survey highlights that econometric developments in predictability mainly address challenges in modeling predictability one or two at a time. It is likely that future research will provide more hybrid approaches that address several if not all forecasting challenges simultaneously, and which can be easily implemented empirically. There is a glimpse of this gradually starting to occur in some of the most recent literature. For example, as mentioned above, Pettenuzzo and Timmermann (2011) address model uncertainty and instability, Zhu and Zhu (2013), address model, regime and parameter uncertainty, and Westerlund and Narayan, (2014a, 2014b) address persistence, endogeneity and heteroskedasticity. However, with the exception of a few models, the performance of hybrid models set against each other as well as against standard approaches is mostly unknown. Future research into robustly evaluating such models is necessary to determine which model provides more
accurate results. Focusing on just the three studies mentioned above, Westerland and Narayan (2014a, 2014b) models might be further extended to cater for time-varying parameters and model uncertainty as in Pettenuzzo and Timmerman (2011) to derive more fine-tuned results. Hence, a mixture of strategies that include data and model features, and model and parameter uncertainty may possibly lead to more superior models. However, for these innovations to occur, it is likely that financial econometric research will need to embrace more of the literature on combination methods, which is a large and growing area of research on its own.

4. Conclusion
The purpose of this paper has been to review econometric developments in the areas of price discovery and price predictability. We have focused on these two topics because they are conceptually linked. Both areas have benefited from similar developments in econometric modelling and, in contrast with areas such as ARCH/GARCH and stochastic volatility, have not been reviewed before.

In terms of the current status of the two literatures, both areas are informed by developments in modelling of unit roots, cointegration and VAR/VECM processes. While this is most evident in the price discovery literature, it is also reflected in the price predictability literature in terms of, for example, addressing persistence and modelling predictability on the presence of instability and uncertainty. Both literatures essentially follow two directions. One is to develop new methods based on relaxing the assumptions of time series models. Examples include allowing for fractionally integrated/cointegrated processes, addressing non-linearities, and accommodating structural breaks and parameter instability. The other is to develop panel-based models to harness the extra power of cross-sectional and time series information. In both literatures, it is difficult to pinpoint the exact state-of-the-art. It very much
depends on the extension/feature one wants to focus on, which, in turn, dictates the assumption(s) one wants to relax. There are few methods dealing with either topic that simultaneously address several issues at once.

In terms of looking forward, this survey suggests several directions. The first flows from evidence that the financial econometric applications in price discovery and price predictability, as well as in other areas of finance, are being fuelled by developments within econometrics itself. Indeed, many of the recent empirical applications in price discovery and price predictability are on the frontier of econometric methods. Given this, further developments in econometric modelling are needed to relax the assumptions and push the boundaries of applications to price discovery and price predictability.

Second, a limitation of both the price discovery and price predictability literatures is that there has largely been a piecemeal approach in which econometric issues have been dealt with on a one-by-one basis. This reflects the manner in which the econometrics literature has evolved. An area for future research, then, is to develop new modelling methods that combine/synthesise recent developments across multiple econometric issues and address more than one issue at once. One obvious way to begin this is through integrating developments in the time series literature into panel-based frameworks.

A third direction is to take a step back and reconsider what insights from recent developments in econometric modelling that challenge stylised facts tell us about long-held beliefs on the apparent empirical regularities in price discovery and price predictability. For example, in the price discovery literature, allowing for fractional
integration and time-varying dynamics is challenging previous findings about whether price discovery occurs in spot or futures markets, as well as the role of fundamental factors in driving prices in financial crises. This opens up the opportunity to develop theory to explain these empirical results.

References


Andrews, D., and W. Ploberger 1994. Optimal tests when a nuisance parameter is present only under the alternative, Econometrica, 62, 1383—1414.


Dolatabadi, S., Narayan, P.K., Nielsen, M.O. and Xu, K. 2014. Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. Keynote address at the Conference on Recent Developments in Financial
Econometrics and Applications, Centre for Economics and Financial Econometrics Research, Deakin University, Australia, December 5-6.


